Revolutionary Genome Sequencing Technologies—The $1,000 Genome

The purpose of this Request for Applications (RFA) is to solicit grant applications to develop novel technologies that will enable extremely low-cost genomic DNA sequencing. Current technologies are able to produce the sequence of a mammalian-sized genome of the desired data quality for $10–50 million; the goal of this initiative is to reduce costs by at least four orders of magnitude, so that a mammalian-sized genome could be sequenced for approximately $1,000. Substantial fundamental research is needed to develop the scientific and technological knowledge underpinning such a major advance. Therefore, it is anticipated that the realization of the goals of this RFA is a long-range effort that is likely to require as much as ten years to achieve. A parallel RFA HG-04-002 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html) solicits grant applications to develop technologies to meet the shorter-term goal of achieving two-orders of magnitude cost reduction in about five years.

The ability to sequence complete genomes and the free dissemination of the sequence data have dramatically changed the nature of biological and biomedical research. Sequence and other genomic data have the potential to lead to remarkable improvement in many facets of human life and society, including the understanding, diagnosis, treatment and prevention of disease; advances in agriculture, environmental science and remediation; and the understanding of evolution and ecological systems.

The ability to sequence many genomes completely has been made possible by the enormous reduction of the cost of sequencing in the past two decades, from tens of dollars per base in the 1980s to a few cents per base today. However, even at current prices, the cost of sequencing a mammalian-sized genome is tens of millions of dollars and, accordingly, we must still be very selective when choosing new genomes to sequence. In particular, we remain very far away from being able to afford to use comprehensive genomic sequence information in individual health care. For this, and many other reasons, the rationale for achieving the ability to sequence entire genomes very inexpensively is very strong.

There are many areas of high priority research to which genomic sequencing at dramatically reduced cost would make vital contributions. 1) Expanded comparative genomic analysis across species, which will yield great insights into the structure and function of the human genome and, consequently, the genetics of human health and disease. Studies to date that have been able to compare small regions of several genomes, and "draft" versions of full genomes, have clearly demonstrated the need for much more complete data sets. While some of the needed data will be obtained over the next two or three years using existing DNA sequencing technology, and while costs will continue their gradual decline, the cost of current approaches to sequence acquisition will continue to limit the amount of useful data that can be produced. 2) Studies of human genetic variation and the application of such information to individual health care, which will also require much cheaper sequencing technology. Today, genetic variation must be assessed by genotyping the relatively few known differences at a relatively small number of loci within the human population. A richer and better characterized catalog of such variable sites is being generated to support more detailed and powerful analyses.

While these methods are, and will become even more, powerful and likely to provide a significant amount of important new information, they are nevertheless only a surrogate for determining the full, contiguous sequence of individual human genomes, and are not as informative as sequencing would be. For example, current genotyping methods are likely to miss rare differences between people at any particular location in the genome and have limited ability to determine long-range information (e.g., genomic rearrangements). Therefore, new methods based on complete genomic sequencing will be needed to use genomic information for individual health care in the most effective manner possible. 3) While the genomes of a few agriculturally important animals and plants have been sequenced, the most informative studies will require comparisons between different individuals, different domesticated breeds and several wild variants of each species. 4) Sequence analysis of microbial communities, many members of which cannot be cultured, would provide a rich source of medically and environmentally useful information. And accurate, rapid sequencing may also be the best approach to microbial monitoring of food and the environment, including rapid detection and mitigation of bioterrorism threats.

Given the broad utility and high importance of dramatically reducing DNA sequencing costs, the National Human Genome Research Institute (NHGRI) is launching two parallel technology development programs. The first has the objective of reducing the cost of producing a high quality sequence of a mammalian-sized genome by two orders of magnitude (see accompanying RFA, HG-04-002). The goal of the second program, described in this RFA, is the development of technology to sequence a genome for a cost that is reduced by four orders of magnitude. For both programs, the cost targets are defined in terms of a mammalian-sized genome, about 3 gigabases (Gb), with a target sequence quality equivalent to, or better than, that of the mouse assembly published in December 2002 [Nature 420:520 (2002)].

The ultimate goal of this program is to obtain technologies that can produce assembled sequence (i.e., *de novo* sequencing). However, an accompanying shorter-term goal is to obtain highly accurate sequence data at the single base level, i.e., without assembly information, that can be overlaid onto a reference sequence for the same organism (i.e., re-sequencing). This could be achieved, for example, with short reads that have no substantial information linking them to other reads. While the sequence product of this kind of technology would lack some important information, such as information about genomic rearrangements, it would nevertheless potentially be available more rapidly and produce data of great value for certain uses in studying disease etiology and in individualized medicine. Therefore, both programs' objectives include a balanced portfolio of projects developing both *de novo* and re-sequencing technologies.

State-of-the-art technology (i.e., fluorescence detection of dideoxynucleotide-terminated DNA extension reactions resolved by capillary array electrophoresis [CAE]) allows the determination of sequence "read" segments approximately 1000 nucleotides long. If all of the DNA in a 2-3 Gb genome were unique, it would be possible to determine the sequence of the entire genome by generating a sufficient number (millions) of randomly-overlapping thousand-base reads and align them by overlaps. However, the human and the majority of other interesting genomes contain a substantial amount of repetitive DNA (short [tens to thousands of nucleotides], nearly or completely identical sequences present in multiple [tens to thousands of] copies). To cope with the complexities of repetitive DNA elements and to assemble the thousand-base reads in the correct long-range order across the genome, current genomic sequencing methods involve a variety of additional strategies, such as the sequencing of both ends of cloned DNA fragments, use of libraries of cloned fragments of different lengths, incorporation of map information, achievement of substantial redundancy (multiple reads of each nucleotide from overlapping fragments) and application of sophisticated assembly algorithms to align and filter the read information.

The "gold standard" for genomic sequencing is 99.99% accuracy (not more than one error per 10,000 nucleotides) with essentially no gaps (http://www.genome.gov/10000923). At present, the final steps in achieving that very high sequence quality cannot be automated and require substantial hand-crafting. However, recent experience suggests that the majority of comparative sequence information can be obtained from automatically generated sequence assemblies that have been variously identified as "high-quality draft" or "comparative grade." Therefore, while the ultimate goal is sequencing technology that produces perfect accuracy, the goal of the current program is to develop technology for producing automatically generated sequence of at least the quality of the mouse draft genome sequence that was published in December 2002 [Nature 420:520 (2002)].

Emerging technologies, collectively characterized as sequencing-by-synthesis or sequencing-by-extension, may be able to achieve large numbers of sequence reads by extending very large numbers of different DNA templates simultaneously, but generally only for a few tens of bases as currently practiced. Even if it is possible to extend these reads to several hundred bases, it will still be necessary to link those reads to achieve long-range sequence contiguity. For some purposes, long-range sequence contiguity may not be required. For example, the re-sequencing of genomes (determination of the DNA sequence for many individuals of a species after a reference sequence for that species has been determined), such as might be used for

medical diagnostic purposes, could be achieved by aligning individual reads on the reference sequence. However, short reads, particularly ones with lower per-base quality, can be very difficult to align given the nature of repetitive DNA and of closely-related gene families in complex genomes. Also, chromosomal rearrangements may be difficult to detect without high quality sequence information bridging the breakpoints with enough sequence to know in which repeat the breakpoint lies. The determination of single nucleotide polymorphisms (SNPs) and their phase (for haplotypes) also requires contiguity of varying length. The ultimate goal and a high priority for the NHGRI's sequencing technology development efforts, as exemplified in these two RFAs, continues to be *de novo*, assembled sequence. However, because of the value of re-sequencing for many future purposes, these RFAs also solicit the development of very inexpensive technology for very high quality re-sequencing (without assembly).

Most investigators interested in reducing DNA sequencing costs anticipate that a few additional two-fold decreases in cost can yet be achieved with the current CAE-based technology, with a realistic lower limit of perhaps $5 million per mammalian-sized genome. However, it is likely that this efficiency will only be achieved in a few very large, well-capitalized, experienced, automated laboratories. To achieve the broadest benefit from DNA sequencing technology for biology and medicine, systems that are not only substantially more efficient but also more usable by the average research laboratory are needed.

One set of current technology development efforts is aimed at increasing parallel sample processing while integrating the sample preparation and analysis steps on a single platform. Thus, in one approach, lithography is used to create a large number of microchannels on a single device and to integrate an efficient sample injector with each separation channel. Chambers for on-chip DNA amplification, cycle sequencing reactions and sample clean-up have been also developed, and experiments to integrate these steps, an approach that effectively places much of the actual process and process control onto the device, are being conducted in several laboratories. Attendant improvements in separation polymers and in fluorescent dyes will facilitate these developments. As these approaches are based largely on the experience of currently successful high-throughput CAE-based methods, they have potential to produce cost savings in the range of several factors of two beyond the CAE-based system itself. They also have the potential to widen the user base for the technology, as the infrastructure and knowledge needed to conduct relatively high-throughput sequencing, or clinical diagnostic sequencing, would be substantially reduced and simplified.

Other approaches to improving sequencing technology involve methods that are independent of the Sanger dideoxynucleotide chain termination reaction or of electrophoretic separation of the termination products. Two methods that were proposed in the early days of the HGP involve the use of mass spectrometry and sequencing by hybridization. Both methods have been pursued, with some limited success for sequencing, but substantial success for other types of DNA analysis. Both continue to hold additional potential utility for sequencing, although certain inherent limitations will need to be overcome.

More recently, additional methodologies have been investigated. These may be classified into two approaches. One is sequencing-by-extension, in which template DNA is elongated stepwise and each extension product is detected. Extension is generally achieved by the action of a polymerase that adds a deoxynucleotide, followed by detection of a fluorescent or chemiluminescent signal; the cycle is then repeated. Modifications of this approach rely on other types of enzymes and detection of hybridization of labeled oligonucleotides. To obtain sufficient throughput, the method is implemented at a high level of multiplexing, e.g., by arraying large numbers of sequencing extension reactions on a surface. A key factor in this general approach is the manner in which the fluorescent signal is generated and the system requirements thus imposed. Depending on the specific approach, challenges of template extension methods include the synthesis of labeled nucleotide analogues; identification of processive polymerases that can incorporate nucleotide analogs with high fidelity; discrimination of fluorescent nucleotides that have been incorporated into the growing chain from those present in the reaction mix (background); distinction of subsequent nucleotide additions from previous ones; accurate enumeration of homopolymer runs (multiple sequential occurrence of the same nucleotide); maintenance of synchrony among the multiple copies of DNA being extended to generate a detectable signal, or achievement of sensitivity that detects extension of individual DNA molecules; and development of fluidics, surface chemistry, and automation to build and run the system. Current efforts to develop such methods have produced, at best, short sequence reads (less than or equal to 100 bases), so a continuing challenge is to extend read length and develop sequence assembly strategies. NHGRI anticipates that the state of the art for this approach is sufficiently advanced that, with additional investment, it may be possible to achieve proof of principle or even early commercialization for genome-scale sequencing within five years. It is anticipated that the cost of genome sequencing with this technology could be reduced by two orders of magnitude from today's costs. It is important to note that sequencing by extension is one prototype for achieving these time and cost goals, but other technological approaches may also be viable. Reaching this goal is the subject of a parallel RFA, HG-04-002 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html).

A second alternative to CAE sequencing seeks to read out the linear sequence of nucleotides without copying the DNA and without incorporating labels, relying instead on extraction of signal from the native DNA nucleotides themselves. The most familiar model for this approach, but almost certainly not the only way to achieve 10,000-fold reduction in sequencing costs, is nanopore sequencing, first introduced in the mid-1990s. Generally, this approach requires a sensor, perhaps comparable in size to the DNA molecule itself, that interacts sequentially with individual nucleotides in a DNA chain and distinguishes between them on the basis of chemical, physical or electrical properties. Optimal implementation of such a method would analyze intact, native genomic DNA molecules isolated from biological, medical or environmental samples without amplification or modification, and would provide very long sequence reads (tens of thousands to millions of bases) rapidly and at sufficiently high redundancy to produce assembled sequence of high quality. NHGRI anticipates that the science and technology needed to reduce sequencing costs by four orders of magnitude, whether by the nanopore or some other approach, will require substantial basic research and development, and may take as long as ten years to achieve. Such a sustained research program is the subject of this RFA.

The goal of research supported under this RFA is to develop new, or improved technology to enable rapid, efficient genomic DNA sequencing. The specific goal is to reduce sequencing costs by at least four orders of magnitude -- $1000 serves as a useful target cost for a mammalian-sized genome because the availability of complete genomic sequences at that cost would revolutionize biological research and medicine. New sensing and detection modalities will likely be needed to achieve these goals. New fabrication technologies may also be required. It is therefore anticipated that proposals responding to this RFA will need to involve fundamental and engineering research conducted by multidisciplinary teams of investigators. The guidance for budget requests accommodates the formation of groups having investigators at several institutions, in cases where that is needed to assemble a team of the appropriate balance, breadth and experience.

The scientific and technical challenges inherent in achieving a 10,000-fold reduction in sequencing costs are clearly daunting. Achieving this goal may require research projects that entail substantial risk. That risk should be balanced by an outstanding scientific and management plan designed to achieve the very high payoff goals of this solicitation.

Although the ultimate goal of this RFA is to develop full-scale sequencing systems, independent research on essential components will also be considered to be responsive. However, it will be important for applicants proposing research on system components or concepts to describe how the knowledge gained as a result of their project would be incorporated into a full system that they might subsequently propose to develop, or that is being developed by other groups. Such independent proposals are an important path for pursuing novel, high risk/high pay-off ideas.

Research conducted under this RFA may include development of the computational tools associated with the technology, e.g., to extract sequence information, including signal processing, and to evaluate sequence quality and assign

confidence scores. It may also address strategies to assemble the sequence from the information being obtained from the technology or by merging the sequence data with information from parallel technology. However, this RFA will not support development of sequence assembly software independent of technology development to obtain the sequence.

The quality of sequence to be generated by the technology is of paramount importance for this solicitation. Two major factors contributing to genomic sequence quality are per-base accuracy and contiguity of the assembly. Much of the utility of comparative sequence information will derive from characterization of sequence variation between species, and between individuals of a species. Therefore, per-base accuracy must be high enough to distinguish polymorphism at the single-nucleotide level (substitutions, insertions, deletions). Experience and resulting policy have established a target accuracy of not more than one error per 10,000 bases. All applications in response to this RFA, whether to develop re-sequencing or *de novo* sequencing technologies, must propose achieving per-base quality at least to this standard.

Assembly information is needed for determining sequence of new genomes, and ultimately also for genomes for which a reference sequence exists, to detect rearrangements, insertions and deletions. Rearrangements are known to cause diseases; knowledge of rearrangement can reveal new biological mechanisms. The phase of single nucleotide polymorphisms to define haplotypes is important in understanding and diagnosing disease. Achieving a high level of sequence contiguity will be essential to achieve the full benefit from the use of sequencing for individualized medicine, e.g., to evaluate genomic contributions to risk for specific diseases and syndromes, and drug responsiveness. Nevertheless, it is recognized that perfect sequence assembly from end to end of each chromosome is unlikely to be achievable with most technologies in a fully automated fashion and without adding considerable cost. Therefore, for the purpose of this solicitation, grant applications proposing technology development for *de novo* sequencing shall describe how they will achieve, for about $1000, a draft-quality assembly that is at least comparable to that represented by the mouse draft sequence produced by December 2002: 7.7-fold coverage, 6.5-fold coverage in Q20 bases, assembled into 225,000 sequence contigs connected by at least two read-pair links into supercontigs [total of 7,418 supercontigs at least 2 kb long], with N50 length for contigs equal to 24.8 kb and for supercontigs equal to 16.9 Mb [Nature 420:520 (2002)].

The grant applications will be evaluated, and funding decisions made, in such a way as to develop a balanced portfolio that has strong potential to develop both robust re-sequencing and *de novo* sequencing technologies. If the estimate that achieving the goal of $1000 *de novo* genome sequencing incorporating substantial assembly information will require about 10 years to achieve is correct, then re-sequencing technologies might be expected to be demonstrated in a shorter time. Grant applications that present a plan to achieve high quality re-sequencing while on the path to high quality *de novo* sequencing will receive high priority.

The major focus of this RFA is on the development of new technologies for detection of nucleotide sequence. However, any new technology will eventually have to be effectively incorporated into the entire sequencing workflow, starting with a biological sample and ending with sequence data of the desired quality, and this issue should be addressed. Given that sample preparation requirements are a function of the detection method and the sample detection method affects the way in which output data are handled, these aspects of the problem are clearly relevant and should be addressed in an appropriate timeframe. However, NHGRI is interested in seeing that the most critical and highest-risk aspects of the project, on which the rest of the project is dependent, are addressed and proven as early as possible.

Practical implementation issues related to workflow and process control for efficient, high quality, high-throughput DNA sequencing should be considered early. Some technology development groups lack practical experience in high throughput sequencing, and in testing of methods and instruments for robust, routine operation. Applicants may therefore wish to include such expertise as they develop their suite of collaborations and capabilities.

The goal of this research is to develop technology to produce sequence from entire genomes. It is conceivable that sequence from selected important regions (e.g., all of the gene regions) could be determined in the near future, using more conventional technologies, at very low cost. However, that is not the purpose of this initiative, and grant applications that propose to meet the cost targets by sequencing only selected regions of a genome will be considered unresponsive.

This RFA will use NIH R21, R21/R33, R01 and P01 award mechanism(s). As an applicant you will be solely responsible for planning, directing, and executing the proposed project.

Applicants may request an R01 or P01 (depending on the organization of the proposed project) if sufficient preliminary data are available to support such an application. A fully integrated management and research plan should use the R01 mechanism. The P01 mechanism should be used if multiple projects under different leadership must proceed in parallel; however, the issue of synergy in a multi-focal effort is of great importance and must be addressed in the application.

Applicants requiring support to demonstrate feasibility may apply for either an R21 pilot/exploratory project or an R21/R33 award, which offers single submission and evaluation of both a feasibility/pilot phase (R21) and an expanded development phase (R33) in one application. The R21/R33 should be used when both quantitative milestones for the feasibility demonstration, and a research plan for the follow-on research, can be presented. The transition from the R21 award to the R33 award will be expedited by administrative review. The R21 alone is appropriate when the possible outcomes of the proposed feasibility study are unclear and it is not possible to propose sufficiently clear-cut and quantitative milestones for administrative evaluation, nor would it be possible to describe the R33 phase of the research in sufficient detail to allow adequate initial review.

This RFA uses just-in-time concepts. It also uses the modular budgeting as well as the non-modular budgeting formats (see http://grants.nih.gov/grants/funding/modular/modular.htm). Specifically, if you are submitting an application with direct costs in each year of $250,000 or less, use the modular budget format. Otherwise follow the instructions for non-modular budget research grant applications. This program does not require cost sharing as defined in the current NIH Grants Policy Statement at http://grants.nih.gov/grants/policy/nihgps_2001/part_i_1.htm. However, cost-sharing is permitted as a component of institutional commitment.

Applications must be prepared using the PHS 398 research grant application instructions and forms (rev. 5/2001). Applications must have a DUN and Bradstreet (D&B) Data Universal Numbering System (DUNS) number as the Universal Identifier when applying for federal grants or cooperative agreements. The DUNS number can be obtained by calling (866) 705-5711 or through the web site at http://www.dunandbradstreet.com/. The DUNS number should be entered on line 11 of the face page of the PHS 398 form. The PHS 398 document is available at http://grants.nih.gov/grants/funding/phs398/phs398.html in an interactive format. For further assistance contact GrantsInfo, 301-435-0714, e-mail: GrantsInfo@nih.gov.

The Center for Scientific Review (CSR) will not accept any application in response to this RFA that is essentially the same as one currently pending initial review, unless the applicant withdraws the pending application. However, when a previously unfunded application, originally submitted as an investigator-initiated application, is to be submitted in response to an RFA, it is to be prepared as a NEW application. That is, the application for the RFA must not include an Introduction describing the changes and improvements made, and the text must not be marked to indicate the changes from the previous unfunded version of the application.

Letters of intent must be received by 14 September 2004. Applications are due by 14 October 2004. The earliest anticipated start date is 1 June 2005.

Contact: Jeffery A. Schloss, Division of Extramural Research, NHGRI, Bldg 31, Rm B2B07, Bethesda, MD 20892-2033 USA, 301-496-7531, fax: 301-480-2770, e-mail: jeff_schloss@nih.gov.

Reference: RFA No. RFA-HG-04-003

Near-term Technology Development for Genome Sequencing

The purpose of this Request for Applications (RFA) is to solicit grant applications to develop novel technologies that will substantially reduce the cost of genomic DNA sequencing. Current technologies are able to produce the sequence of

a mammalian-sized genome of the desired data quality for $10–50 million; the goal of this initiative is to reduce costs by at least two orders of magnitude. It is anticipated that emerging technologies are sufficiently advanced that, with additional investment, it may be possible to achieve proof of principle or even early stage commercialization for genome-scale sequencing within five years. A parallel RFA HG-04-003 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html) solicits grant applications to develop technologies to meet the longer-term goal of achieving four-orders of magnitude cost reduction in about ten years.

The ability to sequence complete genomes and the free dissemination of the sequence data have dramatically changed the nature of biological and biomedical research. Sequence and other genomic data have the potential to lead to remarkable improvement in many facets of human life and society, including the understanding, diagnosis, treatment and prevention of disease; advances in agriculture, environmental science and remediation; and the understanding of evolution and ecological systems.

The ability to sequence many genomes completely has been made possible by the enormous reduction of the cost of sequencing in the past two decades, from tens of dollars per base in the 1980s to a few cents per base today. However, even at current prices, the cost of sequencing a mammalian-sized genome is tens of millions of dollars and, accordingly, we must still be very selective when choosing new genomes to sequence. In particular, we remain very far away from being able to afford to use comprehensive genomic sequence information in individual health care. For this, and many other reasons, the rationale for achieving the ability to sequence entire genomes very inexpensively is very strong.

There are many areas of high priority research to which genomic sequencing at dramatically reduced cost would make vital contributions. 1) Expanded comparative genomic analysis across species, which will yield great insights into the structure and function of the human genome and, consequently, the genetics of human health and disease. Studies to date that have been able to compare small regions of several genomes, and "draft" versions of full genomes, have clearly demonstrated the need for much more complete data sets. While some of the needed data will be obtained over the next two or three years using existing DNA sequencing technology, and while costs will continue their gradual decline, the cost of current approaches to sequence acquisition will continue to limit the amount of useful data that can be produced. 2) Studies of human genetic variation and the application of such information to individual health care, which will also require much cheaper sequencing technology. Today, genetic variation must be assessed by genotyping the relatively few known differences at a relatively small number of loci within the human population. A richer and better characterized catalog of such variable sites is being generated to support more detailed and powerful analyses.

While these methods are, and will become even more, powerful and likely to provide a significant amount of important new information, they are nevertheless only a surrogate for determining the full, contiguous sequence of individual human genomes, and are not as informative as sequencing would be. For example, current genotyping methods are likely to miss rare differences between people at any particular location in the genome and have limited ability to determine long-range information (e.g., genomic rearrangements). Therefore, new methods based on complete genomic sequencing will be needed to use genomic information for individual health care in the most effective manner possible. 3) While the genomes of a few agriculturally important animals and plants have been sequenced, the most informative studies will require comparisons between different individuals, different domesticated breeds and several wild variants of each species. 4) Sequence analysis of microbial communities, many members of which cannot be cultured, would provide a rich source of medically and environmentally useful information. And accurate, rapid sequencing may also be the best approach to microbial monitoring of food and the environment, including rapid detection and mitigation of bioterrorism threats.

Given the broad utility and high importance of dramatically reducing DNA sequencing costs, NHGRI is launching two parallel technology development programs. The first, described in this RFA, has the objective of reducing the cost of producing a high quality sequence of a mammalian-sized genome by two orders of magnitude. The goal of the second program (see accompanying RFA HG-04-003) is the development of technology to sequence a genome for a cost that is reduced by four orders of magnitude. For both programs, the cost targets are defined in terms of a mammalian-sized genome, about 3 gigabases (Gb), with a target sequence quality equivalent to, or better than, that of the mouse assembly published in December 2002 [Nature 420:520 (2002)].

The ultimate goal of this program is to obtain technologies that can produce assembled sequence (i.e., *de novo* sequencing). However, an accompanying shorter-term goal is to obtain highly accurate sequence data at the single base level, i.e., without assembly information, that can be overlaid onto a reference sequence for the same organism (i.e., re-sequencing). This could be achieved, for example, with short reads that have no substantial information linking them to other reads. While the sequence product of this kind of technology would lack some important information, such as information about genomic rearrangements, it would nevertheless potentially be available more rapidly and produce data of great value for certain uses in studying disease etiology and pharmacogenomics, and for comparative genomics between closely-related organisms. Therefore, both programs' objectives include a balanced portfolio of projects developing both *de novo* and re-sequencing technologies.

State-of-the-art technology (i.e., fluorescence detection of dideoxynucleotide-terminated DNA extension reactions resolved by capillary array electrophoresis [CAE]) allows the determination of sequence "read" segments approximately 1000 nucleotides long. If all of the DNA in a 2-3 Gb genome were unique, it would be possible to determine the sequence of the entire genome by generating a sufficient number (millions) of randomly-overlapping thousand-base reads and align them by overlaps. However, the human and the majority of other interesting genomes contain a substantial amount of repetitive DNA (short [tens to thousands of nucleotides], nearly or completely identical sequences present in multiple [tens to thousands of] copies). To cope with the complexities of repetitive DNA elements and to assemble the thousand-base reads in the correct long-range order across the genome, current genomic sequencing methods involve a variety of additional strategies, such as the sequencing of both ends of cloned DNA fragments, use of libraries of cloned fragments of different lengths, incorporation of map information, achievement of substantial redundancy (multiple reads of each nucleotide from overlapping fragments) and application of sophisticated assembly algorithms to align and filter the read information.

The "gold standard" for genomic sequencing is 99.99% accuracy (not more than one error per 10,000 nucleotides) with essentially no gaps (http://www.genome.gov/10000923). At present, the final steps in achieving that very high sequence quality cannot be automated and require substantial hand-crafting. However, recent experience suggests that the majority of comparative sequence information can be obtained from automatically generated sequence assemblies that have been variously identified as "high-quality draft" or "comparative grade." Therefore, while the ultimate goal is sequencing technology that produces perfect accuracy, the goal of the current program is to develop technology for producing automatically generated sequence of at least the quality of the mouse draft genome sequence that was published in December 2002 [Nature 420:520 (2002)].

Emerging technologies, collectively characterized as sequencing-by-synthesis or sequencing-by-extension, may be able to achieve large numbers of sequence reads by extending very large numbers of different DNA templates simultaneously, but generally only for a few tens of bases as currently practiced. Even if it is possible to extend these reads to several hundred bases, it will still be necessary to link those reads to achieve long-range sequence contiguity. For some purposes, long-range sequence contiguity may not be required. For example, the re-sequencing of genomes (determination of the DNA sequence for many individuals of a species after a reference sequence for that species has been determined), such as might be used for medical diagnostic purposes, could be achieved by aligning individual reads on the reference sequence. However, short reads, particularly ones with lower per-base quality, can be very difficult to align given the nature of repetitive DNA and of closely-related gene families in complex genomes. Also, chromosomal rearrangements may be difficult to detect without high quality sequence information bridging the breakpoints with enough sequence to know in which repeat

the breakpoint lies. The determination of single nucleotide polymorphisms (SNPs) and their phase (for haplotypes) also requires contiguity of varying length. The ultimate goal and a high priority for the National Human Genome Research Institute's (NHGRI) sequencing technology development efforts, as exemplified in these two RFAs, continues to be *de novo*, assembled sequence. However, because of the value of re-sequencing for many future purposes, these RFAs also solicit the development of very inexpensive technology for very high quality re-sequencing (without assembly).

Most investigators interested in reducing DNA sequencing costs anticipate that a few additional two-fold decreases in cost can yet be achieved with the current CAE-based technology, with a realistic lower limit of perhaps $5 million per mammalian-sized genome. However, it is likely that this efficiency will only be achieved in a few very large, well-capitalized, experienced, automated laboratories. To achieve the broadest benefit from DNA sequencing technology for biology and medicine, systems that are not only substantially more efficient but also more usable by the average research laboratory are needed.

One set of current technology development efforts is aimed at increasing parallel sample processing while integrating the sample preparation and analysis steps on a single platform. Thus, in one approach, lithography is used to create a large number of microchannels on a single device and to integrate an efficient sample injector with each separation channel. Chambers for on-chip DNA amplification, cycle sequencing reactions and sample clean-up have been also developed, and experiments to integrate these steps, an approach that effectively places much of the actual process and process control onto the device, are being conducted in several laboratories. Attendant improvements in separation polymers and in fluorescent dyes will facilitate these developments. As these approaches are based largely on the experience of currently successful high-throughput CAE-based methods, they have potential to produce cost savings in the range of several factors of two beyond the CAE-based system itself. They also have the potential to widen the user base for the technology, as the infrastructure and knowledge needed to conduct relatively high-throughput sequencing, or clinical diagnostic sequencing, would be substantially reduced and simplified.

Other approaches to improving sequencing technology involve methods that are independent of the Sanger dideoxynucleotide chain termination reaction or of electrophoretic separation of the termination products. Two methods that were proposed in the early days of the HGP involve the use of mass spectrometry and sequencing by hybridization. Both methods have been pursued, with some limited success for sequencing, but substantial success for other types of DNA analysis. Both continue to hold additional potential utility for sequencing, although certain inherent limitations will need to be overcome.

More recently, additional methodologies have been investigated. These may be classified into two approaches. One is sequencing-by-extension, in which template DNA is elongated stepwise and each extension product is detected. Extension is generally achieved by the action of a polymerase that adds a deoxynucleotide, followed by detection of a fluorescent or chemiluminescent signal; the cycle is then repeated. Modifications of this approach rely on other types of enzymes and detection of hybridization of labeled oligonucleotides. To obtain sufficient throughput, the method is implemented at a high level of multiplexing, e.g., by arraying large numbers of sequencing extension reactions on a surface. A key factor in this general approach is the manner in which the fluorescent signal is generated and the system requirements thus imposed. Depending on the specific approach, challenges of template extension methods include the synthesis of labeled nucleotide analogues; identification of processive polymerases that can incorporate nucleotide analogs with high fidelity; discrimination of fluorescent nucleotides that have been incorporated into the growing chain from those present in the reaction mix (background); distinction of subsequent nucleotide additions from previous ones; accurate enumeration of homopolymer runs (multiple sequential occurrence of the same nucleotide); maintenance of synchrony among the multiple copies of DNA being extended to generate a detectable signal, or achievement of sensitivity that detects extension of individual DNA molecules; and development of fluidics, surface chemistry, and automation to build and run the system. Current efforts to develop such methods have produced, at best, short sequence reads (less than or equal to 100 bases), so a continuing challenge is to extend read length and develop sequence assembly strategies. NHGRI anticipates that the state of the art for this approach is sufficiently advanced that, with additional investment, it may be possible to achieve proof of principle or even early commercialization for genome-scale sequencing within five years. It is anticipated that the cost of genome sequencing with this technology could be reduced by two orders of magnitude from today's costs. It is important to note that sequencing by extension is one prototype for achieving these time and cost goals, but other technological approaches may also be viable. Developing technology with which to reduce the cost of genome sequencing by 100-fold is the subject of this RFA.

A second alternative to CAE sequencing seeks to read out the linear sequence of nucleotides without copying the DNA and without incorporating labels, relying instead on extraction of signal from the native DNA nucleotides themselves. The most familiar model for this approach, but almost certainly not the only way to achieve 10,000-fold reduction in sequencing costs, is nanopore sequencing, first introduced in the mid-1990s. Generally, this approach requires a sensor, perhaps comparable in size to the DNA molecule itself, that interacts sequentially with individual nucleotides in a DNA chain and distinguishes between them on the basis of chemical, physical or electrical properties. Optimal implementation of such a method would analyze intact, native genomic DNA molecules isolated from biological, medical or environmental samples without amplification or modification, and would provide very long sequence reads (tens of thousands to millions of bases) rapidly and at sufficiently high redundancy to produce assembled sequence of high quality. NHGRI anticipates that the science and technology needed to reduce sequencing costs by four orders of magnitude, whether by the nanopore or some other approach, will require substantial basic research and development, and may take as long as ten years to achieve. Reaching this goal is the subject of a parallel RFA, HG-04-003 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html).

The goal of research supported under this RFA is to develop or improve technology to enable rapid, efficient genomic DNA sequencing. The specific goal is to reduce sequencing costs by at least two orders of magnitude—$100,000 serves as a useful target cost for a mammalian-sized genome because the availability of complete genomic sequences at that cost would revolutionize biological research and medicine. While not in a cost range that would enable the use of sequencing in individualized medicine, such technology would permit the sequencing of many genomes for a small fraction of current costs. A 100-fold cost reduction would make possible extensive studies of human variation for disease gene studies, substantially expanded comparative genomics to understand the human genome, and many other studies relevant to the National Institutes of Health (NIH), other federal agencies and the private sector. Entirely new lines of investigation would be enabled by making "large-scale sequencing" accessible to the diverse interests of many research laboratories and companies.

Many projects aimed at next-generation DNA sequencing technologies require substantial advances in a combination of fields such as signal detection, enzymology, chemistry, engineering, bioinformatics, etc. It is therefore anticipated that research programs responding to this RFA will involve multidisciplinary teams of investigators. The guidance for budget requests accommodates the formation of groups having investigators at several institutions, in cases where that is needed to assemble a team of the appropriate balance, breadth and experience.

The scientific and technical challenges inherent in achieving a 100-fold reduction in sequencing costs are considerable. Achieving this goal may require research projects that entail substantial risk. That risk should be balanced by an outstanding scientific and management plan designed to achieve the very high payoff goals of this solicitation.

Although the ultimate goal of this RFA is to develop full-scale sequencing systems, independent research on essential components will also be considered to be responsive. However, it will be important for applicants proposing research on system components or concepts to describe how the knowledge gained as a result of their project would be incorporated into a full system that they might subsequently propose to develop,

or that is being developed by other groups. Such independent proposals are an important path for pursuing novel, high risk/high pay-off ideas.

Research conducted under this RFA may include development of the computational tools associated with the technology, e.g., to extract sequence information, including signal processing, and to evaluate sequence quality and assign confidence scores. It may also address strategies to assemble the sequence from the information being obtained from the technology or by merging the sequence data with information from parallel technology. However, this RFA will not support development of sequence assembly software independent of technology development to obtain the sequence.

The quality of sequence to be generated by the technology is of paramount importance for this solicitation. Two major factors contributing to genomic sequence quality are per-base accuracy and contiguity of the assembly. Much of the utility of comparative sequence information will derive from characterization of sequence variation between species, and between individuals of a species. Therefore, per-base accuracy must be high enough to distinguish polymorphism at the single-nucleotide level (substitutions, insertions, deletions). Experience and resulting policy have established a target accuracy of not more than one error per 10,000 bases. All applications in response to this RFA, whether to develop re-sequencing or *de novo* sequencing technologies, must propose achieving per-base quality at least to this standard.

Assembly information is needed for determining sequence of new genomes, and ultimately also for genomes for which a reference sequence exists, to detect rearrangements, insertions and deletions. Rearrangements are known to cause diseases; knowledge of rearrangement can reveal new biological mechanisms. The phase of single nucleotide polymorphisms to define haplotypes is important in understanding and diagnosing disease. Achieving a high level of sequence contiguity will be essential to achieve the full benefit from the use of sequencing for individualized medicine, e.g., to evaluate genomic contributions to risk for specific diseases and syndromes, and drug responsiveness. Nevertheless, it is recognized that perfect sequence assembly from end to end of each chromosome is unlikely to be achievable with most technologies in a fully automated fashion and without adding considerable cost.

Therefore, for the purpose of this solicitation, grant applications proposing technology development for *de novo* sequencing shall describe how they will achieve, for about $1000, a draft-quality assembly that is at least comparable to that represented by the mouse draft sequence produced by December 2002: 7.7-fold coverage, 6.5-fold coverage in Q20 bases, assembled into 225,000 sequence contigs connected by at least two read-pair links into supercontigs [total of 7,418 supercontigs at least 2 kb long], with N50 length for contigs equal to 24.8 kb and for supercontigs equal to 16.9 Mb [Nature 420:520 (2002)].

The grant applications will be evaluated, and funding decisions made, in such a way as to develop a balanced portfolio that has strong potential to develop both robust re-sequencing and *de novo* sequencing technologies. If the estimate that achieving the goal of 100-fold reduction in cost for genome sequencing incorporating substantial assembly information will require about 5 years to achieve is correct, then re-sequencing technologies might be expected to be demonstrated in a shorter time. Grant applications that present a plan to achieve high quality re-sequencing while on the path to high quality *de novo* sequencing will receive high priority.

The major focus of this RFA is on the development of new technologies for detection of nucleotide sequence. However, any new technology will eventually have to be effectively incorporated into the entire sequencing workflow, starting with a biological sample and ending with sequence data of the desired quality, and this issue should be addressed. Given that sample preparation requirements are a function of the detection method and the sample detection method affects the way in which output data are handled, these aspects of the problem are clearly relevant and should be addressed in an appropriate timeframe. However, NHGRI is interested in seeing that the most critical and highest-risk aspects of the project, on which the rest of the project is dependent, are addressed and proven as early as possible.

NHGRI anticipates that successful projects funded through this RFA may be sufficiently advanced as to be approaching early stages of commercialization within about five years. Therefore, practical implementation issues related to workflow and process control for efficient, high quality, high-throughput DNA sequencing should be considered early. Some technology development groups lack practical experience in high throughput sequencing, and in testing of methods and instruments for robust, routine operation. Applicants may therefore wish to include such expertise as they develop their suite of collaborations and capabilities.

The goal of this research is to develop technology to produce sequence from entire genomes. It is conceivable that sequence from selected important regions (e.g., all of the gene regions) could be determined in the near future, using more conventional technologies, at very low cost. However, that is not the purpose of this initiative, and grant applications that propose to meet the cost targets by sequencing only selected regions of a genome will be considered unresponsive.

This RFA will use NIH R21, R21/R33, R01 and P01 award mechanism(s). As an applicant you will be solely responsible for planning, directing, and executing the proposed project.

Applicants may request an R01 or P01 (depending on the organization of the proposed project) if sufficient preliminary data are available to support such an application. A fully integrated management and research plan should use the R01 mechanism. The P01 mechanism should be used if multiple projects under different leadership must proceed in parallel; however, the issue of synergy in a multi-focal effort is of great importance and must be addressed in the application.

Applicants requiring support to demonstrate feasibility may apply for either an R21 pilot/exploratory project or an R21/R33 award, which offers single submission and evaluation of both a feasibility/pilot phase (R21) and an expanded development phase (R33) in one application. The R21/R33 should be used when both quantitative milestones for the feasibility demonstration, and a research plan for the follow-on research, can be presented. The transition from the R21 award to the R33 award will be expedited by administrative review. The R21 alone is appropriate when the possible outcomes of the proposed feasibility study are unclear and it is not possible to propose sufficiently clear-cut and quantitative milestones for administrative evaluation, nor would it be possible to describe the R33 phase of the research in sufficient detail to allow adequate initial review.

This RFA uses just-in-time concepts. It also uses the modular budgeting as well as the non-modular budgeting formats (see http://grants.nih.gov/grants/funding/modular/modular.htm). Specifically, if you are submitting an application with direct costs in each year of $250,000 or less, use the modular budget format. Otherwise follow the instructions for non-modular budget research grant applications. This program does not require cost sharing as defined in the current NIH Grants Policy Statement at http://grants.nih.gov/grants/policy/nihgps_2001/part_i_1.htm. However, cost-sharing is permitted as a component of institutional commitment.

Applications must be prepared using the PHS 398 research grant application instructions and forms (rev. 5/2001). Applications must have a DUN and Bradstreet (D&B) Data Universal Numbering System (DUNS) number as the Universal Identifier when applying for Federal grants or cooperative agreements. The DUNS number can be obtained by calling (866) 705-5711 or through the web site at http://www.dunandbradstreet.com/. The DUNS number should be entered on line 11 of the face page of the PHS 398 form. The PHS 398 document is available at http://grants.nih.gov/grants/funding/phs398/phs398.html in an interactive format. For further assistance contact GrantsInfo, 301-435-0714, e-mail: GrantsInfo@nih.gov.

The Center for Scientific Review (CSR) will not accept any application in response to this RFA that is essentially the same as one currently pending initial review, unless the applicant withdraws the pending application. However, when a previously unfunded application, originally submitted as an investigator-initiated application, is to be submitted in response to an RFA, it is to be prepared as a NEW application. That is, the application for the RFA must not include an Introduction describing the changes and improvements made, and the text must not be marked to indicate the changes from the previous unfunded version of the application.

Letters of intent must be received by 14 September 2004. Applications are due 14 October 2004. The earliest anticipated start date: 1 June 2005.

Contact: Jeffery A. Schloss, Division of Extramural Research, NHGRI, Bldg 31, Rm B2B07, Bethesda, MD 20892-2033 USA, 301-496-7531, fax: 301-480-2770, e-mail: jeff_schloss@nih.gov.

Reference: RFA No. RFA-HG-04-003

In Vivo Cellular and Molecular Imaging Centers (ICMICS)

The Cancer Imaging Program, Division of Cancer Diagnosis and Treatment of the National Cancer Institute (NCI), invites applications for new or competing P50 Research Center Grants for *In vivo* Cellular and Molecular Imaging Centers (ICMICs). This initiative is designed to capitalize on the extraordinary opportunity for molecular imaging to have an impact on the diagnosis and treatment of cancer patients non-invasively and quantitatively. Molecular imaging technologies can provide valuable laboratory tools for the interrogation of biological pathways relevant to cancer, as well as to provide imaging agents and technologies that will be directly utilized in the clinic. The five-year P50 ICMIC grants described in this PAR are designed to bring together interdisciplinary scientific teams to lead the nation in cutting-edge cancer molecular imaging research with clinical relevance, provide unique core facilities to support oncology imaging research, provide flexibility to respond to exciting pilot research opportunities, and provide interdisciplinary career development opportunities for investigators new to the field of molecular cancer imaging. The P50 mechanism will promote coordination, interrelationships and scientific synergy among the research components and resources, leading to a highly integrated imaging center.

The field of molecular imaging has made significant advances in recent years. The formation of multidisciplinary research teams has stimulated and streamlined cancer imaging research from inception to use in patient care. The P50 ICMIC structure allows mechanistic flexibility for each Institution to capitalize on its own unique scientific strengths, and to define the structure and research objectives that create the most synergistic and creative scientific interactions. In general, an ICMIC will provide researchers with the following critical resources:

The ICMICs will provide an organizational structure specifically designed to facilitate multidisciplinary interactions among investigators focused on the ultimate goal of discovering, developing and translating molecular imaging technologies that will have eventual impact in the clinic. This structure will provide researchers with access to a concentrated pool of expertise in a wide range of disciplines. The structure of the ICMIC will be designed to provide investigators with the means of conducting multidisciplinary research in a highly collaborative atmosphere, and consistent access to expertise with minimal wasted time and effort. Personnel may be scientists from a variety of fields including, but not limited to: imaging sciences, chemistry, radiopharmaceutical chemistry, cell and molecular biology, pathology, pharmacology, computational sciences, and biomedical engineering. Other specialists in fields such as MRI physics, immunology, or neuroscience, for example, may also be involved. Most importantly, ICMIC personnel must demonstrate an eagerness to collaborate outside of their own disciplines. The nature of these interactions will be determined by the applicants, and emphasis will be placed on establishing creative, productive, and synergistic interactions with eventual clinical impact.

The ICMICs will provide funding for a minimum of three Research Components. Research Components will apply multidisciplinary approaches to molecular imaging. Individual research projects will be structured in order to maximize appropriate scientific interaction between the projects, and coordinated utilization of the Specialized Resources (see below). Each Research Component will be similar in size and scope to a typical R01 or subproject of a P01, and will be expected to meet the same standards of preliminary data in support of the hypotheses.

The ICMICs will provide Specialized Resource Facilities and Services. A barrier to productive scientific interaction is the lack of available facilities for cross-disciplinary experiments. Demands on equipment, resources, and reagents in every scientific area are extremely high, and this demand prohibits ready access to investigators interested in expanding their studies into new areas of research. The establishment of Specialized Resources dedicated to ICMIC-related research will provide this access. The Specialized Resource(s) will be determined by the requirements of the Institution, the defined scientific goals of the Research Components of the ICMIC, and budgetary limits. Prioritization of the research projects supported through ICMIC Specialized Resources will be an essential function of the ICMIC's leadership, and the mechanism to be employed for prioritization must be delineated by the applicants. Resource facilities may be utilized by active members of the ICMIC and will also be available to investigators supported through Developmental Funds (see below).

ICMICs will provide Developmental Funds for feasibility testing of new projects. A high priority of each ICMIC will be the identification and support of pilot projects that identify and stimulate interdisciplinary projects that will take full advantage of emerging research opportunities. The selection of projects will be through a review process established by the ICMIC's leadership. The portfolio of ongoing projects in any given Program is expected to be extremely dynamic. This fund is not to be used to support traditional, ongoing projects that could readily be supported through R01s. It is not appropriate for projects that utilize single areas of expertise or to support the continuation of previously funded research projects, and Developmental Projects may not be supported for more than two years. Necessary equipment should be provided through the appropriate Specialized Resource. These projects are to be monitored closely by the ICMIC leadership. Investigators working on projects supported through the Development Fund must understand that they will be expected to compete for independent R01 funding when the projects become sufficiently mature. Alternatively, if it becomes obvious that the project will not provide the expected results, a plan should be in place for terminating a development project.

ICMICs will provide career development opportunities for new and established investigators. Current graduate programs are generally focused on single disciplines and may be inadequate to train the needed cadre of inter-disciplinary imaging scientists. The ICMICs will provide support for a limited number of pre-and post-doctoral trainees in a program to be defined by the applicants. Career development opportunities through the ICMIC will be expected to be highly cross-disciplinary.

This PAR will use the NIH P50 Specialized Centers Grant Mechanism. As an applicant, you will be solely responsible for planning, directing, and executing the proposed project. The total project period for a P50 application submitted in response to this PAR may not exceed five years. The total costs requested for a new or competing renewal P50 ICMIC application may not exceed a maximum of $2,000,000 per year. The NCI anticipates awarding two new or competing P50 ICMICs each year.

This PAR uses just-in-time concepts. It also uses the non-modular budgeting formats. Follow the instructions for non-modular budget research grant applications. This program does not require cost sharing as defined in the current NIH Grants Policy Statement at http://grants.nih.gov/grants/policy/nihgps_2003/NIHGPS_Part2.htm#Toc54600040.

Applications must be prepared using the PHS 398 research grant application instructions and forms (rev. 5/2001). Applications must have a Dun and Bradstreet (D&B) Data Universal Numbering System (DUNS) number as the Universal Identifier when applying for federal grants or cooperative agreements. The DUNS number can be obtained by calling (866) 705-5711 or through the web site at http://www.dunandbradstreet.com/. The DUNS number should be entered on line 11 of the face page of the PHS 398 form. The PHS 398 document is available at http://grants.nih.gov/grants/funding/phs398/phs398.html in an interactive format. For further assistance contact GrantsInfo, 301-435-0714, e-mail: GrantsInfo@nih.gov.

Applications hand-delivered by individuals to the NCI will no longer be accepted. This policy does not apply to courier deliveries (i.e. FEDEX, UPS, DHL, etc.) See http://grants.nih.gov/grants/guide/notice-files/NOT-CA-02-002.html for more information. This policy is similar to and consistent with the policy for applications addressed to Centers for Scientific Review as published in the NIH Guide Notice at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-012.html.

Applications must be received on or before the receipt date(s) listed on the first page of this PA. The CSR will not accept any application in response to this PAR that is essentially the same as one currently pending initial review unless the applicant withdraws the pending application. The CSR will not accept any application that is essentially the same as one already reviewed. This does not preclude the submission of a substantial

revision of an unfunded version of an application already reviewed, but such application must include an Introduction addressing the previous critique. 398 research grant application instructions (rev. 5/2001) will be assessed.

Letters of intent must be received by 22 June 2004 and 21 June 2005. Applications must be received by 22 July 2004 and 21 July 2005. The earliest anticipated start dates are April 2005 and April 2006.

Contact: Anne E. Menkens, Cancer Imaging Program, NCI, 6130 Executive Blvd, EPN Rm 6068, Bethesda, MD 20892-8329 USA, (Rockville, MD 20852 for express/courier service), 301-496-9531, fax: 301-480-3507, e-mail: am187k@nih.gov.

Reference: PA No. PAR-04-069

Pharmacogenetics Research Network and Knowledge Base

The purpose of this RFA is to solicit applications for an open re-competition of the Pharmacogenetics Research Network and Knowledge Base (http://www.nigms.nih.gov/pharmacogenetics). This is a network of multidisciplinary, collaborative groups of investigators that contribute their data to the publicly available knowledge base PharmGKB, which is an open research tool accessible to all scientists.

The research groups in the network have interests across a range of biological processes: drug metabolism, small molecule transport, target receptors, and biological pathways involved in the drug treatment of cardiovascular diseases, asthma, cancer, and depression; other areas are welcome consistent with the interests of the funding institutes. The groups are collecting comprehensive, integrative information about specific proteins and gene families important to the field of pharmacogenetics. Some groups are using a genotype-to-phenotype approach starting with the detection of all possible variants, while other groups are employing a phenotype-to-genotype approach beginning with well-characterized clinical samples. All investigations are converging on the association of single nucleotide polymorphisms (SNPs) and haplotypes with drug responses. The results are confirmed by studies of the mechanistic and clinical consequences of the molecular changes. The database groups in the network are working towards the goal of creating a centralized public knowledge base. PharmGKB (http://www.pharmgkb.org) is designed to categorize four types of phenotype information—functional assays, pharmacokinetics, pharmacodynamics, and clinical outcomes—correlated with genotype information. The knowledge base uses standardized drug, disease, and genetic vocabularies and is linked to existing databases.

The plans are to continue funding this network as a series of cooperative groups conducting studies to address a wide variety of common research problems in pharmacogenetics. This initiative will further emphasize development of the PharmGKB knowledge base; it is envisioned as an information resource that will be useful to the entire pharmacogenetics research community to enable future hypothesis-driven research. This competition is open to both new and renewal research and database groups.

Pharmacogenetics can be defined as the influence of human genetic variation on drug responses. It has long been known from family studies that variations found in enzymes of drug clearance have profound effects on the efficacy and duration of drug action, sometimes with significant adverse consequences. Genetic variations in drug metabolizing enzymes can lead to the excessive build-up of a drug with a narrow therapeutic index (e.g., thiopurine methyl-transferase and 6-mercaptopurine), or the lack of a therapeutic effect where metabolic activation is required (e.g., cytochrome P450 2D6 and codeine). Likewise, studies have shown that variations in target receptors can lead to a lack of beneficial effects of a drug, for example by increased desensitization (e.g., beta-2 adrenoreceptor and albuterol).

Another mechanism impacting drug efficacy is altered binding kinetics (e.g., serotonin 1B receptors and fluoxetine). Recent studies have shown that genetic variants can be linked to the susceptibility and progression of disease as well as to a response to a drug treatment (e.g., cholesterol ester transfer protein, atherosclerosis, and statins; or apolipoprotein E, Alzheimer's disease, and tacrine). There are multiple genetic mechanisms, including alterations in transcript stability, splice sites, or promotor binding regions, all of which can alter expression levels. The impact of these changes on functional protein levels such as reduced amounts or stability, or compromised enzymatic function, requires further study. Furthermore, how this fits into protein-protein interactions (e.g., coupling to second messengers) and biological pathways (e.g., redundant, competing, or complementary routes of clearance or signaling) needs to be understood in order to predict clinical consequences.

With advances in genomic technology, large-scale accumulation of information on drug pathways (sometimes called pharmacogenomics) is possible. These profiling studies can be DNA-based, transcript-based, or protein-based. Both pharmacogenetics and pharmacogenomics studies are of interest under this solicitation. It is essential to completely understand the significance of genetic variation at the molecular level, and the implications of the diverse genetic contexts present in different human populations. The incidence of SNPs (singly and in combinations of haplotypes) and gene duplication or deletion events must be interpreted correctly to associate genetic variation with the prediction of drug effects, and this may require development of new analytical tools. Population-based studies that examine the interactions between genetic predisposition for disease and the genetic factors determining medication responses are also of interest for this initiative.

Ultimately, both a mechanistic understanding and robust statistical validation of putative pharmacogenetics effects are sought, and the translation to clinical impact is highly desirable. The goal of the field is to be able to predict the effects of a medication in an individual based upon his/her genome, but much research must be performed before that is possible in a comprehensive manner. Accurate descriptions of drug response phenotypes are challenging and difficult, and further research is required to define these phenotypes. The Pharmacogenetics Research Network is intended to address this need to acquire basic research results and store the information in a knowledge base, which will lead to a more complete understanding of drug actions, clinical translation of the information, and future drug development.

The Pharmacogenetics Research Network will continue to be comprised of a series of multidisciplinary research and database groups, each of which is performing state-of-the-art studies in pharmacogenetics, either independently or in conjunction with other network groups.

While pursuing the highest quality research studies, each network group must agree to meet the following expectations: 1) to further develop the knowledge base, PharmGKB, which is a database with accurate and detailed definitions of pharmacogenetic phenotypes linked to genotypes; 2) to advance the research field, by defining common goals and needs, and contributing to solving problems of the field through discussions and workshops; 3) to produce and share resources, such as biological reagents, and experimental and computational tools, to be disseminated rapidly and with minimal restrictions; and 4) to communicate with scientists both within and outside the network, and to foster translation and application of this knowledge. These requirements are further detailed below, and are included in the specific review criteria.

A research group should be organized around a unifying theme, for example, a family of proteins with which drugs interact, a set of drug pathways leading to the site of action, or drug treatments for a particular disease. The group should be comprised of a multidisciplinary team of investigators, minimally including personnel with backgrounds in cellular/molecular pharmacology, genetics/genomics, and clinical expertise. Individuals from the fields of pharmacology, pharmaceutics, physiology, genetics, genomics, clinical medicine, medicinal chemistry, epidemiology, statistics, bioinformatics, and computational biology may be incorporated and must demonstrate that they can work together. This research team should propose current, cutting-edge pharmacogenetics studies. They should be "driven by the science" to produce the highest quality research results for deposition into PharmGKB and for publication. The research groups will be responsible for serving as interactive resources for the developers of PharmGKB in their self-described areas.

Applications should not simply be proposed as a series of projects from all investigators working in pharmacogenetics at an institution. Careful thought should be given to the definition of a research group's goals, and the steps to be taken to accomplish those goals. The best core or project teams to accomplish the research goals should be assembled; applications that cross multiple institutions are acceptable. An application should discuss how existing databases were used to design and approach the solution of a pharmacogenetic problem, and how PharmGKB can

better serve its users in the future. The assembled group must justify their choice of a research area as the most appropriate, demonstrate their study design and power, and employ state-of-the-art technical approaches, including statistics and analyses. The selected research problem in pharmacogenetics could be conceived starting with the identification of all possible variants (a genotype-to-phenotype approach) or beginning with well-characterized patient materials (a phenotype-to-genotype approach). The applicant group should state the advantages and disadvantages of the approach chosen, and where convergence is expected with other studies ongoing in the field.

Correct and complete descriptions of phenotypes and association with genotypes form the core organizing principle underlying the Pharmacogenetics Research Network. The research groups being funded are required to produce meaningful data sets suitable to populate PharmGKB. Scientifically valid research questions should be constructed to yield data that contribute to advancing the understanding in the field, and that are appropriate for deposition into the knowledge base. The types of data deposits that are expected should be described in detail, along with the time frame for their submission. Both human and animal data, as well as nonmammalian systems, will be accepted. Where animals or cell lines or model organisms are being examined, they should be justified as the appropriate reference models, consistent with the goal of identifying and interpreting human genetic drug response variants.

Research groups should address how the pharmacogenetic researchers outside of the network can be positively impacted. Useful sample sets should be offered to established repositories (e.g., the National Institute of General Medical Sciences [NIGMS] Human Genetic Cell Repository at the Coriell Institute, http://locus.umdnj.edu/nigms/) for immortalization and distribution. Useful reagents (e.g., antibodies, primers) should be made easily available. Software tools should be shared freely whenever possible.

Current papers representative of the research field being studied should be deposited by the research groups into the community submissions project in PharmGKB. Evidence of these steps taken will attest to the desire of the research group to serve in a scientific network and to share their findings with the scientific community, and should be presented in the application.

A database group applying to continue PharmGKB should present a plan to further develop the knowledge base as a research resource that will store, organize, present, and integrate pharmacogenetic knowledge. PharmGKB must display a variety of data types: genetic variants, haplotypes, population frequencies, summary statistics, oligonucleotide and cDNA microarray data, molecular and functional screening assays, pharmacokinetic data, pharmacodynamic data, and, where appropriate, clinical data demonstrating the consequences of genetic variation. It should have in place user-friendly methods to accept these data deposits of diverse forms and sizes. In all cases, the data should be described using the standard nomenclature of the respective fields. The knowledge base should have reciprocal links to other established databases, such as GenBank, dbSNP, PDB, etc.

The knowledge base should describe gene-protein-drug-disease relationships, with each object layer completely represented. Relationships between these different data types should be displayed visually, and reflect the opinions and agreement of researchers working in these fields. Raw data should be stored wherever possible, so that PharmGKB can be mined to learn of new correlations. This is intended to be a hypothesis-generating tool. Moreover, data should also be summarized and interpreted so that the information in the knowledge base is accessible to all scientist-users. Given the long history of the field of pharmacogenetics, there should be a current and complete literature archive linked to complete publications wherever possible. Existing high value data sets outside of the network research groups should be sought to populate PharmGKB, to ensure complete and even-handed representation across the field of pharmacogenetics. Methods to establish credit and provide practical scientific incentives for submitters should be proposed.

Applications to continue the knowledge Base PharmGKB should reflect the current status of the project, and describe how the design aspects, implementation, and maintenance will be continued or improved upon. Careful attention should be paid to issues of curation, and delineating who has the responsibility to format, abstract, and check the different kinds of data sets for completeness and accuracy. Comparisons should be made to other successful ongoing database efforts. Future major design directions should be presented and discussed, with prototypes. Discussion of accomplishments, challenges, and obstacles should be provided, and/or external observations and alternative strategies on how to overcome problem areas. There should be evidence of the practical ability to work with the research groups in the network. If a new database group is funded, copies of the existing datasets and data tables will be provided at the time of award, according to the prior negotiated terms and conditions regarding future portability.

Taken together, the research and database groups of the Pharmacogenetics Research Network and Knowledge Base should encompass a range of ongoing studies and original data on pharmacologically important genes, proteins, and pathways. This will be accomplished by funding a balanced series of research groups that are studying different gene families, drug treatments, and diseases of significance to human health. The scope of the Pharmacogenetics Research Network will likely continue to include enzymes of drug metabolism, small molecule transporters, and target receptors and pathways involved in drug treatment of cardiovascular diseases, asthma, cancer, and depression, and may broaden somewhat in reflection of the participating NIH institutes' interests.

This network will be continued as a trans-NIH effort; the institutes' specific interests: NIGMS is interested in studies identifying robust, statistically valid correlations between pharmacogenetic responses (phenotypes) and genetic variation (genotypes, haplotypes) using state-of-the-art approaches and technologies, and in the deposition of this knowledge into a database designed to be accessible by the entire research community.

The National Cancer Institute (NCI) is interested in projects that can potentially lead to meaningful improvements in clinical and survival endpoints, and in studies of genetic variability in human populations that may influence risk of preneoplastic conditions or primary and secondary malignancies after exposure to medications, including cancer therapies.

The National Heart, Lung, and Blood Institute (NHLBI) is interested in studies of the role of genetic polymorphisms and their functional consequences in modulating treatment responses in heart, lung, blood, and sleep diseases.

The National Human Genome Research Institute (NHGRI) supports research on how databases represent phenotypes, particularly related to genetic variation, and encourages the use and extension of standardized ontologies, as well as rapid data release.

The National Institute on Drug Abuse (NIDA) is interested in the influence of genetic variation on metabolic, homeostatic, neurocognitive, and physiological responses to abused drugs, as well as the safety and efficacy of drugs used for the treatment of addiction, dependence, and withdrawal, and in drug-drug interactions (e.g., antiretrovirals and drugs of abuse).

The National Institute of Environmental Health Sciences (NIEHS) is interested in identifying the response genes that are important to understanding genetic susceptibility to environmental exposures (see the Environmental Genome Project at http://www.niehs.nih.gov/envgenom/home.htm).

The National Library of Medicine (NLM) is interested in knowledge representation and the design and management of databases with medical data.

The Office of Research on Women's Health (ORWH) is interested in evaluating the importance of gender differences in genetic polymorphisms of proteins important in the pharmacokinetics and pharmacodynamics of drugs and drug reactions, and the role of hormones and other factors.

This RFA will use the NIH U01 award mechanism. The applicant is solely responsible for planning, directing, and executing the proposed project. The RFA is a one-time solicitation. The anticipated award date is on or after 1 July 2005. Applications that are not funded in the competition described in this RFA may be resubmitted as NEW investigator-initiated applications using the standard receipt dates for NEW applications described in the instructions to the PHS 398 application.

This RFA uses just-in-time concepts. It uses the nonmodular budgeting formats. Follow the instructions for nonmodular budget research grant applications and submit the detailed categorical budget information on the PHS 398 form. This program does not require cost sharing

as defined in the current NIH Grants Policy Statement at http://grants.nih.gov/grants/policy/nihgps_2001/part_i_1.htm.

The NIH U01 is a cooperative agreement award mechanism. In the cooperative agreement mechanism, the Principal Investigator retains the primary responsibility and dominant role for planning, directing, and executing the proposed project, with NIH staff being substantially involved as a partner with the Principal Investigator as described under the section "Cooperative Agreement Terms and Conditions of Award." NIH makes no commitment to continue the cooperative agreement programs beyond the initially awarded period of performance.

Attendance at two Steering Committee meetings per year is required. These will likely rotate between the East and West coasts and central United States. Travel funds should be requested for this purpose for the Principal Investigator and for one to two other Observers. A plan for depositing data into PharmGKB is required. See the current submission methods at http://www.pharmgkb.org/submit/index.jsp.

This satisfies the NIH requirement for sharing research data for applications greater than $500,000 direct costs in any year of the proposed research. Funds should be requested to support individuals capable of submitting data to PharmGKB.

A letter should be included in the application, stating that the applicant research group members have read all of the existing policies of the Pharmacogenetics Research Network (http://pharmgkb.org/home/policies/index.jsp). The letter should indicate that the group members will adhere to each of the policies and will contribute to the development of future policies that will guide the network's actions.

Applications must be prepared using the PHS 398 research grant application instructions and forms (rev. 5/2001). Applications must have a DUN and Bradstreet (D&B) Data Universal Numbering System (DUNS) number as the Universal Identifier when applying for federal grants or cooperative agreements. The DUNS number can be obtained by calling (866) 705-5711 or through the web site at http://www.dunandbradstreet.com/. The DUNS number should be entered on line 11 of the face page of the PHS 398 form. The PHS 398 document is available at http://grants.nih.gov/grants/funding/phs398/phs398.html in an interactive format. For further assistance contact GrantsInfo, 301-435-0714, e-mail: GrantsInfo@nih.gov.

Using the RFA label: The RFA label available in the PHS 398 (rev. 5/2001) application form must be affixed to the bottom of the face page of the application. Type the RFA number on the label. Failure to use this label could result in delayed processing of the application such that it may not reach the review committee in time for review. In addition, the RFA title and number must be typed on line 2 of the face page of the application form and the YES box must be marked. The RFA label is also available at: http://grants.nih.gov/grants/funding/phs398/labels.pdf.

The Center for Scientific Review (CSR) will not accept any application in response to this RFA that is essentially the same as one currently pending initial review, unless the applicant withdraws the pending application. However, when a previously unfunded application, originally submitted as an investigator-initiated application, is to be submitted in response to an RFA, it is to be prepared as a NEW application. That is, the application for the RFA must not include an Introduction describing the changes and improvements made, and the text must not be marked to indicate the changes from the previous unfunded version of the application.

Letters of intent must be received by 19 July 2004. Applications are due 19 August 2004. The earliest anticipated start date is 1 July 2005.

Contact: Rochelle M. Long, Pharmacology, Physiology, and Biological Chemistry Division, NIGMS, NIH, Bldg 45, Rm 2AS.49G, MSC 6200, Bethesda, MD 20892-6200 USA, 301-594-1926, fax: 301-480-2802, e-mail: longr@nigms.nih.gov; Richard A. Anderson, Genetics and Developmental Biology Division, NIGMS, NIH, Bldg 45, Rm 2AS.25B, MSC 6200, Bethesda, MD 20892-6200 USA, 301-594-0943, fax: 301-480-2228, e-mail: andersor@nigms.nih.gov; Ken Kobayashi, Cancer Therapy Evaluation Program, NCI, NIH, 6130 Executive Blvd, Ste 7131, MSC 7426, Rockville, MD 20852-4907 USA, 301-496-1196, fax: 301-402-0428; e-mail: kobayashik@ctep.nci.nih.gov; J. Fernando Arena, Division of Cancer Control and Population Sciences, NCI, NIH, 6130 Executive Blvd, Executive Plz N, MSC 7395, Rm 5104, Rockville, MD 20852-4907 USA, 301-594-5868, fax: 301-402-4279, e-mail: arenaj@mail.nih.gov; Susan Banks-Schlegel, Division of Lung Diseases, NHLBI, NIH, Rockledge Two, Rm 10220, 6701 Rockledge Dr, MSC 7952, Bethesda, MD 20992-0001 USA, 301-435-0202, fax: 301-480-3557, e-mail: schleges@nih.gov; Dina Paltoo, Division of Heart and Vascular Diseases, NHLBI, NIH, Rockledge Two, Rm 9180, 6701 Rockledge Dr, MSC 7940, Bethesda, MD 20892-0001 USA, 301-435-1802, fax: 301-480-1336, e-mail: paltood@mail.nih.gov; Lisa D. Brooks, Genetic Variation Program, NHGRI, NIH, 31 Center Dr, Rm B2B07, Bethesda, MD 20892-2033 USA, 301-435-5544, fax: 301-480-2770, e-mail: lisa_brooks@nih.gov; Joni L. Rutter, Division of Neuroscience and Behavioral Research, NIDA, NIH, 6001 Executive Blvd, Rm 5227, MSC 9555, Bethesda, MD 20892-9555 USA, 301-435-0298, fax: 301-594-6043, e-mail: jrutter@mail.nih.gov; Kimberly Gray, Division of Extramural Research and Training, NIEHS, NIH, 111 T.W. Alexander Drive, PO Box 12233, MD EC-21, Research Triangle Park, NC 27709 USA, 919-541-0293, fax: 919-316-4606, e-mail: kg89o@nih.gov; Milton Corn, Extramural Programs, NLM, NIH, 6705 Rockledge Dr, Bldg 1, Ste 301, Bethesda, MD 20892-0001 USA, 301-496-4621, fax: 301-402-2952, e-mail: cornm@mail.nlm.nih.gov; Lisa Begg, Research Programs, ORWH, OD, NIH, One Center Dr, Rm 201, MSC 0161, Bethesda, MD 20892-0001 USA, 301-496-7853, fax: 301-402-1798, e-mail: beggl@od.nih.gov.

Reference: RFA No. RFA-GM-04-002

Strategic Partnering to Evaluate Cancer Signatures

The purpose of this initiative is to build on recent demonstrations that molecular signatures correlate with important clinical parameters in cancer. The National Cancer Institute (NCI) invites investigators to form strategic partnerships that will bring together the multi-disciplinary expertise and resources needed to determine how the information derived from comprehensive molecular analyses can be used to improve patient care and ultimately, patient outcomes. Applicants are asked to propose evaluation of potential clinical usefulness of molecular signatures already developed using a variety of molecular analysis technologies including DNA, RNA or protein-based technologies.

Molecular signatures have been able, in retrospective studies, to identify subgroups of patients whose tumors are histopathologically the same but who have different clinical outcomes. The challenge is to translate the information in these molecular signatures into tools that can be used in clinical decision-making. To meet this challenge, signatures must be confirmed in independent studies. Critical elements of signatures that correlate most strongly with the clinical endpoint of interest must be identified and confirmed. Robust assays feasible for use in the clinical setting must be developed and validated. This iterative process of signature refinement and confirmation and assay refinement requires diverse scientific expertise and access to significant patient and tissue resources.

This initiative is an open competition that will provide the cancer research community the opportunity to establish collaborations focused on the translation of promising molecular profiles toward clinical application.

NCI will continue the policy of requiring public release in a timely fashion of the rich data sets generated during these projects. Access to these data sets will benefit the entire cancer research community. This initiative will help ensure that the NCI goal of eliminating the suffering and death from cancer by 2015 is met.

The projects funded by this RFA are intended to exploit the successes of the many research projects applying comprehensive molecular analysis in cancer. Comprehensive molecular technologies have been demonstrated to provide a snapshot of the biological state of a tumor. The ability of molecular profiles to provide useful clinical information is now being demonstrated in many projects throughout the cancer research community and needs to be evaluated further. Projects are discovering molecular signatures by analysis of gene expression at the RNA level, gene expression following protein translation, gene mutations, DNA deletions, DNA amplifications, epigenetic changes of DNA and post-translational modification of proteins. The challenge is to move beyond the initial discovery of potentially useful profiles, to decide what subset of the elements in the profiles needs to be measured, to confirm that the profiles are robust and can be reproducibly measured and to evaluate the clinical utility of the profiles.

This RFA is open to all interested, qualified investigators. The initiative is intended to support projects carrying out the extensive research

needed to bridge the gap between discovery of molecular profiles and their integration into clinical decision-making. Applicants should propose projects that address clinical issues or needs in a specific cancer or a closely related set of cancers or in a group of patients whose cancers have related molecular alterations. Collaborations must be established to provide all of the expertise and clinical resources required to achieve proposed project goals. It is anticipated that these will be multi-institutional projects involving investigators with expertise in technology development and application, cancer biology, oncology, pathology, clinical cancer research, biostatistics, bioinformatics and, possibly, biomedical imaging.

Applicants must propose projects that build on previously identified molecular profiles. Applications proposing only profile discovery or technology development projects will not be considered responsive to this RFA. The proposed studies should be designed to confirm and refine signatures that have been demonstrated to provide information that is potentially useful clinically and that may be used to aid in making clinical decisions. Applicants may propose to define critical components in the signature, to confirm that the selected components continue to provide the desired clinical information and to develop robust assays for measuring those components. They may continue to develop and/or modify analytical technologies and algorithms for data analysis required to meet the goals of the proposed projects.

Applicants must establish the collaborations necessary to bring together the expertise needed for the project. Successful completion of the project will require expertise in analytical technologies, cancer biology, oncology, pathology, clinical cancer research, biostatistics, bioinformatics and possibly biomedical imaging.

Applicants must describe the clinical question(s) or need(s) they plan to address. The clinical questions posed should address a well-defined clinical need in one or a closely related set of tumors or in a group of patients whose cancers have related molecular alterations. Examples of questions of interest may include, but are not limited to: risk of progression in early stage disease; prognosis at the time of diagnosis; identification of subsets within a tumor stage or grade where there is known heterogeneity in clinical behavior including differential response to standard therapies and/or radiation response; and selection of appropriate patients for or prediction of response to selected or targeted therapies. Applicants to this initiative should not propose projects addressing early detection of cancer in asymptomatic or high-risk populations or risk of progression of pre-malignant lesions.

Applicants may propose the use of a variety of analytical platforms(s). Applicants may propose to evaluate signatures the have previously been identified using analytical technologies such as, but not limited to, gene expression microarrays, SAGE, multiplex PCR or any of a large number of protein analysis technologies. Genomic analysis technologies such as array CGH, comprehensive mutational analysis technologies, SNP analysis and analysis of epigenetic events are also appropriate. Applicants must demonstrate that they have experience with the analytical technologies that will be used in the project and demonstrate that the technologies can be used for analysis of standard pathological specimens. Applicants are encouraged, but not required, to propose the use of multiple analytical strategies. For example, projects may be proposed to analyze gene expression in both frozen tissue and paraffin-embedded tissue, to analyze gene expression at both the RNA and protein level or to analyze both epigenetic alterations and gene expression. The integration of data to build clinically useful profiles that can be measured reproducibly in a clinical setting must be the focus of the project, no matter which technologies or analytic platforms are proposed.

Applicants must justify the numbers of specimens to be analyzed based on appropriate statistical designs for the proposed studies. Applicants must have established collaborations to ensure availability of the clinical materials required. The availability of tissue resources with appropriate clinical annotation is critical to the successful completion of the projects. Experience has demonstrated that the dimensionality of the molecular profiling data requires the analysis of hundreds, not tens, of specimens to get statistically significant results. Applicants may propose to obtain tissues from a previous collection or prospectively, as long as the specific aims proposed can be accomplished within the period of the grant award. Demonstrated access to the requisite tissues will be critical to the successful review of the application. It is recognized that tissue needs may change as the projects are carried out. NCI staff will work with investigators to help identify additional tissue resources needed to meet project goals.

Applicants should request sufficient resources to ensure that they will be able to collect, manage and analyze the data generated. Applicants must address issues related to obtaining, managing and controlling the quality of the clinical data needed for specimen annotation. Continued development of strategies to more effectively address issues of data management and analysis will be an inter-project cooperative activity of the funded projects.

Applicants should request sufficient resources for their bioinformatics staff to be able to provide an appropriate interface with the NCI Center for Bioinformatics (NCICB). Sharing of the data between projects where appropriate and public release of data after publication will be a requirement for this initiative. Gene expression data will be shared through the NCICB Gene Expression Database using the Gene Expression Data Portal and database. Proteomics data and other types of data can be shared through the NCICB site as the capabilities of the site are expanded. They may also be shared through investigators'websites or on other publicly available websites.

The confirmation, refinement and evaluation of clinically useful molecular profiles and the development of robust clinical assays are the primary goals of this initiative. Clinical utility of the signatures and performance of the clinical assays in the context of their intended clinical use must be validated before they can be integrated into clinical practice. Final validation of the profiles in a clinical trial setting is beyond the scope of this initiative. However, it is anticipated that some of the projects may be ready to move profiles into clinical trials as early as the midpoint of the project period. NCI staff will facilitate collaborations between the projects funded on this initiative and other clinical resources and clinical trials activities supported by NCI including: the clinical cooperative groups; the Program for the Assessment of Clinical Cancer Tests (PACCT); the SPORE programs and the NCI Cancer Centers.

This RFA will use the Natiional Institutes of Health (NIH) cooperative agreement (U01) award mechanism. As an applicant you will be solely responsible for planning, directing, and executing the proposed project. This RFA is a one-time solicitation. The anticipated award date is 1 April 2005. Applications that are not funded in the competition described in this RFA may be resubmitted as NEW investigator-initiated applications using the standard receipt dates for NEW applications described in the instructions to the PHS 398 application.

This RFA uses just-in-time concepts. It also uses the non-modular budgeting formats. Follow the instructions for non-modular budget research grant applications. This program does not require cost sharing as defined in the current NIH Grants Policy Statement at http://grants.nih.gov/grants/policy/nihgps_2001/part_i_1.htm.

The NIH (U01) is a cooperative agreement award mechanism. In the cooperative agreement mechanism, the Principal Investigator retains the primary responsibility and dominant role for planning, directing, and executing the proposed project, with NIH staff being substantially involved as a partner with the Principal Investigator, as described under the section "Cooperative Agreement Terms and Conditions of Award." At this time, it is not known if this RFA will be reissued.

In order to ensure maximum progress in the projects funded by this initiative and to maximize progress toward the NCI 2015 goals, several special activities will be required of the funded investigators. An annual meeting of all funded investigators will be held to share progress and research insights that may benefit all of the projects.

The annual scientific meeting will be initiated after the first year of funding. One or more other focused meetings will be held each year to address arising issues or to take advantage of special scientific opportunities. Applicants should request travel funds in their budgets for key personnel to attend two meetings per year.

The funded investigators will be asked to work together on issues common to all funded projects. Although each applicant will propose an independent project, all applicants are expected to face many of the same challenges and will benefit from the experiences of and interactions with the other funded investigators. The interactions of the funded groups will be overseen by a

Steering Committee made up of two investigators, the PI and one additional investigator, from each funded project and appropriate NCI staff members. A Steering Committee organizing meeting will be held shortly after funding is initiated. The Steering Committee will focus on common problems and issues, especially issues of data management and analysis. Applicants should state in their applications their commitment to participating on the Steering Committee and in interactions among the funded groups.

When proposed studies involve collection of human samples, specimens and/or clinical data, investigators should consider the issues raised and guidance provided in the NIH Brochure entitled "Research on Human Specimens: Are You Conducting Research Using Human Subjects?" (http://www-cdp.ims.nci.nih.gov/policy.html) and in the OHRP Guidance on Repositories, Tissue Storage Activities and Data Banks (http://ohrp.osophs.dhhs.gov/g-topicstest.htm) to ensure appropriate protection of human subjects in research.

Applicants must describe how they intend to meet the NIH policies for sharing of data or why data sharing is not possible. In this regard, attention is drawn to the NIH Final Statement on Sharing Research Data (http://grants.nih.gov/grants/policy/data_sharing/index.htm and http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html), which was published in the NIH Guide on 26 February 2003 ("Data Sharing Guidelines"). This is an extension of NIH policy on sharing research resources, and reaffirms NIH support for the concept of data sharing. The new policy becomes effective with the 1 October 2003 receipt date for applications or proposals to NIH. Investigators submitting an NIH application will be required to include a plan for data sharing or to state why data sharing is not possible. The statement required by this section should be prepared with reference to the provision below for Awardee Rights and Responsibilities within Terms and Conditions of Award.

Intellectual property (IP) issues continue to provide challenges to the establishment of complex, collaborative projects. This issue is being addressed and managed in different ways by many different projects supported by NCI. The policy of the NIH is to make available to the public the results and accomplishments of the activities that it funds. NIH recognizes that certain research activities may result in inventions and that grantees are entitled to protect such inventions through patenting and licensing activities in accordance with the Bayh-Dole Act, 35 USC § 200 et seq. and the implementing regulations, 37 CFR Part 401 ("Bayh-Dole Act"). To address the interest in assuring that research resources are accessible, NCI requires applicants who respond to this RFA to submit a plan (1) for sharing the unique research resources generated through the grant; and (2) addressing how they will exercise IP rights, should any be generated through this grant, while making such research resources available to the broader scientific community. The sharing of research resources and IP plans must make unique research resources readily available for research purposes to qualified individuals within the scientific community in accordance with the NIH Grants Policy Statement (http://grants.nih.gov/grants/policy/nihgps_2001/) and the Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice, December 1999 (http://ott.od.nih.gov/NewPages/RTguide_final.html and http://ott.od.nih.gov/NewPages/64FR72090.pdf) ("NIH Research Tools Guidelines"). These documents also define terms, parties, responsibilities, prescribe the order of disposition of rights, prescribe a chronology of reporting requirements, and delineate the basis for and extent of government actions to retain rights. Patent rights clauses may be found at 37 CFR Part 401.14 and are accessible from the Interagency Edison webpage (http://www.iedison.gov/).

If applicant investigators plan to collaborate with third parties, the research tools sharing plan must explain how such collaborations will not restrict their ability to share research materials produced with NIH funding. All applicants will be expected to have addressed IP issues with their proposed collaborators before submitting their applications and to have documented the status of their arrangements by providing a copy of a signed agreement, a signed Memorandum of Understanding between collaborating institutions or a letter of collaboration countersigned by all relevant parties that describes the IP issues and provides applicants with all rights necessary to perform activities required by the research plans. Successful applicants are expected to have resolved any outstanding IP issues before funding is awarded. It is anticipated that successful applicants may subcontract with third party for profit institutions to perform certain aspects of the research plans described in their applications. Successful applicants will be expected to ensure that they obtain sufficient rights in such subcontracts to enable them to meet their obligations under the NIH Research Tools Guidelines and the Data Sharing Guidelines, both of which are extensions of the distribution of unique research resources policy contained in the NIH Grants Policy Statement (page 11-62). In drafting these subcontracts, grantees will want to give some thought to the relative rights and responsibilities of the parties, particularly with respect to the dissemination and use of raw data, results and analyses.

The NCI Technology Transfer Branch staff works with NCI funded investigators on IP issues and has developed strategies for sharing IP. Staff of the NCI Technology Transfer Branch will provide their expertise as needed to the investigators funded under this initiative and their technology transfer and grants and contracts administration offices. Addressing IP issues will be a component of the review of the projects.

As discussed earlier, grantees will be required to publicly release data to the cancer research community through the NCI Center for Bioinformatics web site or through other appropriate public websites. Applicants should commit to the public release of data and request funds in their budgets to support bioinformatics staff interactions with the NCICB staff.

Applications must be prepared using the PHS 398 research grant application instructions and forms (rev. 5/2001). Applications must have a DUN and Bradstreet (D&B) Data Universal Numbering System (DUNS) number as the Universal Identifier when applying for federal grants or cooperative agreements. The DUNS number can be obtained by calling (866) 705-5711 or through the web site at http://www.dunandbradstreet.com/. The DUNS number should be entered on line 11 of the face page of the PHS 398 form. The PHS 398 document is available at http://grants.nih.gov/grants/funding/phs398/phs398.html in an interactive format. For further assistance contact GrantsInfo, 301-435-0714, e-mail: GrantsInfo@nih.gov.

Using the RFA label: The RFA label available in the PHS 398 (rev. 5/2001) application form must be affixed to the bottom of the face page of the application. Type the RFA number on the label. Failure to use this label could result in delayed processing of the application such that it may not reach the review committee in time for review. In addition, the RFA title and number must be typed on line 2 of the face page of the application form and the YES box must be marked. The RFA label is also available at: http://grants.nih.gov/grants/funding/phs398/labels.pdf.

Applications hand-delivered by individuals to the NCI will no longer be accepted. This policy does not apply to courier deliveries (i.e. FEDEX, UPS, DHL, etc.) See http://grants.nih.gov/grants/guide/notice-files/NOT-CA-02-002.html for more information. This policy is similar to and consistent with the policy for applications addressed to the Center for Scientific Review (CSR) as published in the NIH Guide Notice at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-012.html.

The CSR will not accept any application in response to this RFA that is essentially the same as one currently pending initial review, unless the applicant withdraws the pending application. However, when a previously unfunded application, originally submitted as an investigator-initiated application, is to be submitted in response to an RFA, it is to be prepared as a NEW application. That is the application for the RFA must not include an Introduction describing the changes and improvements made, and the text must not be marked to indicate the changes from the previous unfunded version of the application.

Letters of intent must be received by 22 June 2004. Applications are due 22 July 2004. The earliest anticipated start date is 1 April 2005.

Contact: James W. Jacobson, Division of Cancer Treatment and Diagnosis, NCI, 6130 Executive Blvd, EPN Rm 6035A, Bethesda, MD 20892-0001 USA, 301-402-4185, fax: 301-402-7819, e-mail: jacobsoj@mail.nih.gov; Tracy Lugo, Division of Cancer Treatment and Diagnosis, NCI, 6130 Executive Blvd, EPN Rm 6035A, Bethesda, MD 20892 USA, 301-496-1591, fax: 301-402-7819, e-mail: lugot@mail.nih.gov.

Reference: RFA No. RFA-CA-04-015