

**SUMMARY STATEMENT
(Privileged Communication)**

Jeffery Schloss PH.D.
301-435-5538
schlossj@mail.nih.gov

Release Date: 12/23/2003

CHURCH, GEORGE M PHD
HARVARD MEDICAL SCHOOL
DEPARTMENT OF GENETICS
200 LONGWOOD AVE, WA BLDG, RM 513
BOSTON, MA 02115

Application Number: 1 P50 HG003170-01
Formerly: 1P20HG003170-01

Review Group: GNOM-G
Genome Research Review Committee

Meeting Date: 11/17/2003
Council: JAN 2004
Requested Start: 05/01/2004

RFA/PA: PAR02-021
PCC: X7JS

Project Title: Molecular and Genomic Imaging Center

SRG Action: Priority Score: 143

Human Subjects: E4-Human subjects involved - Exemption #4 designated

Animal Subjects: 30-Animals involved - no SRG comments or concerns noted

Children: 4A-Child representation unknown, scientifically acceptable
Clinical Research - not NIH-defined Phase III Trial

Project Year	Direct Costs Requested	Estimated Total Cost
1	2,459,155	3,452,833
2	2,269,231	3,186,166
3	2,209,156	3,101,816
4	2,344,881	3,292,384
5	2,431,250	3,413,652
TOTAL	11,713,673	16,446,851

ADMINISTRATIVE BUDGET NOTE: The budget shown is the requested budget and has not been adjusted to reflect any recommendations made by reviewers. If an award is planned, the costs will be calculated by Institute grants management staff based on the recommendations outlined below in the COMMITTEE BUDGET RECOMMENDATIONS section.

1P50HG003170-01 Church, G.

RESUME AND SUMMARY OF DISCUSSION: This application was submitted in response to an NHGRI program announcement to establish Centers of Excellence in Genomic Science (CEGS). This P50 program is designed to support multi-investigator, interdisciplinary teams to develop innovative genomic approaches to address biological problems. In addition to research activities, the Center is expected to have plans for training and facilitating the interaction of investigators from different disciplines in order to promote genomic science at its institution.

This CEGS application proposes to establish a Center that is focused around the development of a technology called polony or polymerase colony for identifying and characterizing variation at the level of genomes and transcriptomes. A polony is a DNA colony that has been created in an acrylamide gel matrix by PCR amplification of a single nucleic acid molecule. In a highly parallel process, millions of spatially separated polonies can be generated and visualized on a single microscope slide. The Center will continue to develop the technology and apply it to a variety of applications in nucleic acid analysis, including in situ DNA sequencing, comprehensive quantitative measurements of mRNA, single molecule profiling of alternatively spliced exons, direct molecular haplotyping over long genomic distances, and characterization of DNA sequence and gene expression profiles from single cells. Computational algorithms and systems modeling for nucleic acid analysis will also be developed.

The Genome Research Review Committee expressed considerable enthusiasm for this application, convinced that the polony method will be an enabling technology that will have many important applications. They cited the compelling rationale outlined for pursuing the development of the technology for each of the applications proposed and were convinced that other applications will be identified in the future. The committee was optimistic that the work will have substantial impact in a number of these areas and thus the Center was given high marks for significance. Preliminary evidence has been published supporting the feasibility of the approach for several of the applications, which gave the committee added confidence that large dividends will come from an investment in this Center. Although considerable work has been done in developing the approach, both the polony technology itself and its application to a broad spectrum of analyses was considered highly innovative.

The committee was excited about the application of the technology for studying splice variants and the unique technical advance and increased capability this method appears to represent. Feasibility of the approach has been established, and a study using the polony technique for examining alternative exons has been published. One particular advantage may be in finding rare alternative forms that go undetected by other approaches. The committee was also excited about use of the polony technique for direct molecular haplotyping over long genomic distances, and again there has been published evidence supporting the feasibility of approach. There will be a continuing need for this capability and this method could have quite a substantial impact. The committee was less certain about the technology's application to in situ sequencing. The committee concluded that it was unlikely that the investigators would be able to achieve the proposed 500 base read lengths from the current read length of 5-10 bases. The factors that may limit read lengths were not discussed in the application, and although the investigators indicate they will systematically address this issue, the lack of details about how this would be accomplished was disappointing. The committee suggested that one major limiting factor might be the acrylamide matrix in which the PCR reactions take place. Its use imposes inherent limitations that significantly limit the achievable read lengths. There was no evidence that the investigators have investigated this, or are seriously considering alternatives. The applicants suggest that the 100 base lengths reported by pyrosequencing systems gives them some confidence that they can substantially increase their read lengths, but the committee was skeptical, again largely because of the inherent limitations imposed by the acrylamide matrix. At the same time, there were members who noted that even a more modest increase in read length capability would be useful. Much basic work needs to be done to establish the capability of the polony technique for measuring DNA and RNA

in single cells, and while success is not assured, the committee was convinced that the potential payoff is high and fully justifies the effort proposed. Additional technical concerns are detailed in the individual critiques, but overall there was confidence that this group would address the weaknesses. In the end, these concerns did not substantially detract from the committee's overall enthusiasm for the application.

The principal investigator (PI) is a major strength of the application. He is an outstanding scientist and most importantly for the Center, has the stature to direct the overall activities and make the hard decisions about the directions the project will take. The research team provides an excellent core of expertise that gave the committee confidence that the Center will be able to develop and push the technology in many directions. The Center brings together a multidisciplinary mix of biologist, bioengineers, and computational scientists. The committee suggested that expertise in gel matrices might have to be added to encourage the exploration of alternatives to polyacrylamide that may help extend to the read lengths.

The management plan for Center seems well worked out between the investigators at the various institutions. A key collaboration is with Dr. Mitra at Washington University, who was instrumental in developing the polony technique while working with Dr. Church at Harvard. The activities will be coordinated by frequent visits and an advisory board comprised of distinguished scientists will provide an annual review of the Center's progress.

The committee was very impressed by the record and commitment of the PI regarding the dissemination of data and materials as evidenced by the protocols and other information available from the PI's web site.

The general training plan was well developed and there was little question that the environment would provide a rich training experience. The Minority Action Plan was folded into the general training plan, and exhibited both strengths and weaknesses. These are discussed in greater detail in a separate section at the end of this summary statement.

In summary, the committee rated this application as outstanding, citing the high significance of the Center goals, the high level of innovation of the underlying technology, the creative applications to which the technology is being tested, the multidisciplinary nature of the research, and confidence that the work coming out of the Center will improve analytical capabilities in a variety of applications and have a dramatic impact on these fields.

The comments in the CRITIQUE section were prepared by the reviewers assigned to this application and are provided without significant modification or editing by staff. The RESUME AND SUMMARY OF DISCUSSION section documents the final outcome of the evaluation by reviewers and is the basis for the assigned priority score.

DESCRIPTION: (provided by applicant) We propose here the Molecular and Genomic Imaging Center (MGIC) in response to a biomedical-community-wide need for flexible, cost-effective, high-resolution technology to identify and characterize variation in biological systems at the level of genomes and transcriptomes. We plan to help meet this need by developing the polymerase colony, or polony, technology. Polonies represent a highly parallel method of nucleic acid analysis that is realistic, close-at-hand, modular, and versatile. The primary mission of the MGIC is to efficiently integrate a diverse set of contributions from technology developers into a robust platform that can be smoothly disseminated to a variety of users with specialized clinical and biological interests. These are the main aims: (1) Highly parallel fluorescent in situ sequencing. (2) Single molecule profiling of the transcriptome, in particular of neural differentiation and of mammalian alternative splicing. (3) Direct molecular haplotyping and long-range sequence connectivity. (4) Characterization of DNA & RNA from

single cells, in particular characterizing asymmetric cell division in mammalian stem cells. (5) Computational algorithms and systems modeling addressing combinatorial and spatial patterns in nucleic acid analysis. (6) An ELSI component focuses on issues of translation of technology to clinical applications and challenges to the concepts of anonymity.

CRITIQUE 1:

This application builds on a fundamental technical development called 'colonies' where a single DNA fragment is amplified within a gel matrix so that all the daughter fragments are captured locally. Each resulting 'colony' represents a single molecule and is a substrate for sequencing or hybridization assays. The power of the method is that it could allow a facile rescue of single molecules for further analysis.

SIGNIFICANCE: This technology could provide an extremely powerful and generalized mechanism for all kinds of DNA studies. The underlying idea of separating single DNA molecules and then assaying them separately is a good one. Even if the technology has to be carried out in specialized centers, it would yield critical biological information. If you can imagine that one day this might be practiced routinely in individual laboratories with kit-reagents, then this would be a wonderful addition to current technologies.

APPROACH: This multi site application will pursue several aims. First they will try to improve the basic colony methodology by attending to issues like improving matrix and PCR conditions and chemistry and trying to improve automation of the method. The current state of the art is to be able to make 5-8 base reads on some amplified colonies. Aim 1 is to raise this to 20-500 base pairs. The other parts of Aim 1, e.g. matrix improvement and automation, are coupled to this underlying ambition.

In this part of the application there are not a lot of specific innovations proposed. There is ample good discussion of the potential and limitations of the various types of sequencing that can be performed and there is good reason to believe that this outstanding group can carry out good experiments for optimization of the available alternatives. However, aside from some alternative reducing agents for the removal of fluorescent tags stuck on by the disulphide linkages, there are not any actually new methods proposed for the sequencing.

There is some discussion of pyrosequencing and the length of read obtained there is cited as a reason that the reads here could get that long. This logic is flawed as the pyrosequencing methods benefit from the ability of the reagents for the light generation to freely migrate to where they are needed. Those reactions are generally carried out in liquid phase and therefore can be cycled quite quickly. The effect of slowly diffusing in these reagents is not discussed, but it is hard to imagine that a pyrosequencing type approach is best for this methods.

It is extremely likely that this group will advance this basic technology. There is however some places where it just hard to see from this application exactly how they will do that. The issues they will likely wrestle with as part of their attempts to improve the colonies are the seeding density, PCR reagent concentrations, and all that basic stuff. Precise tuning if these will be the key. This kind of improvement will be essential if the subsequent aims are to be pursued.

As part of aim 1 specific disease locus associations will be screened for. In these experiments the association is known but the experiment is novel in that it aims to find these alleles for the first time in these samples. In this reviewers opinion this is not yet the right experiment – for now these investigators should simply be looking for specific bases known to be present in different samples and situations.

In Aim 2 expression profiling will be pursued. To simplify this and to overcome one of the ongoing problems of 'sequencing by synthesis' the homopolymeric tracts will essentially be ignored in this study (i.e. GAAAGTTG will be read as GAGTG). This has adequate coding potential to still recognize tags in relatively modest reads. Alternatively, probe hybridization will be used to detect the specific polonies from a particular probe. To avoid lots of probe hybridizations a pooling scheme is proposed, although not much detail is given. These methods will be applied to studies of differentiation in mouse ES and brain cells. A positive inclusion is the aim to study small numbers of cells.

Overall however, most of these proposed expression experiments in Aim 2A are a weakness in this proposal. While there may be a place for those kinds of experiments in the polony world, these are not the best variations on the methods to aim for. The real strength of the polonies is that they recover information without first needing to know what are the specific sequences. In particular, resorting to specific probe hybridizations places such a limitation on the data set that will be derived that it would be better if other methods were used in the first place.

Aim 2B is a much stronger part of the proposal. Here alternative splicing will be measured along the lines of what has been achieved in the murine CD44 locus. Here there is good preliminary data and a demonstrated understanding of limitations and expectations of the approach. These experiments could very well enhance our understanding of alternative splicing and gene variation.

The first part of Aim 3 is to carry out direct molecular haplotyping. This is a strength of the polony approach and exactly the kind of application that can work the best for it. Several targeted experiments aim to take advantage of the fact that unbroken chromosomes will lie close to one another in the matrix and be suitable to seed co-amplification of multiple priming sites.

Other parts of Aim 3 are to carry out mate pair sequencing and to stretch DNA (akin to fiber fish) in order to amplify along the length of the molecule. These efforts are less positive in the application since they require reasonably long read lengths, although if successful the developments would be very interesting.

In Aim 4 there are *in situ* polonies and these will be explored in order to study quite complex biological phenomena – strand retention preference in non-malignant cells and other mechanisms of asymmetric cell growth and differentiation. These are really very interesting applications and a good sign of the depth to which this outstanding group is able to think and plan their experiments.

In AIM 5 it is planned to build some software to manage all this data and to perform some molecular mathematical modeling. It is certainly true that the absence of any proper software to manage these data is a big factor in the project – this will have to be rectified in order for the utility of the basic innovation to be realized and for the program to move ahead. This seems to have two broad phases: first, the group aims to be able to catalogue and manipulate data and for this some real software engineering will be needed. Next some mathematics will be required. It is not clear how the software engineering will be accomplished. Although the group has adequate skills to generate software it does not have a record yet in making robust and extensive software suites.

Lastly there is an ELSI section that has 3 lectures to discuss ELSI issues.

Overall the approach is sound. The main criticism that could be leveled at this application is that it overly emphasizes general gains that can be made from the success of any single molecule method, but does not really make many specific suggestions about how the exact polony approach can be made more efficient. While experiments that are aimed at optimizations and tuning are not necessarily the most exciting to read about, it is precisely this kind of work that will make this technology a dominant one in the future.

In the discussion of the optimizations there is not much distinction about the kinds of end-user that might take up with the technology. It would have been encouraging to see that area grown more and some specific goals established around the general theme of exportability of the methods. It is clear that many of the developments have the potential to make the deployment easier but this is not really stated or emphasized as a particular goal.

Another issue that is not discussed at all is the operational cycle time for the methods. The applicants provided some update in this area and it appears 35 minutes at least is needed for 'non detection' cycles. In some of the multistep detections that are proposed this adds up to quite a long time – a 500 base read with a protocol that has several steps in it would be prohibitive. This is why some procedures are to be more favored than others as the methods are in development. In particular the single step identification methods (e.g. SNP detection) are real winners in this methodology. The application would have been improved by some better up front discussion of the cycle time issue.

Overall there are many strengths and weaknesses in the technical aspects of this proposal. The biggest weaknesses are in the attempts to exploit the methods in ways that do not play to its strengths. Also the possibility that the more ambitious sequence lengths is unlikely. Nevertheless the methodology has real potential and therefore worthy of considerable enthusiasm.

MANAGEMENT: This is a multi site/multi lab project, with involvement of Labs at Harvard, the University of Delaware and Wash U. There is also a collaboration at the MIT Genome Center. Most of the work is to be carried out in the Church lab and that is where most of the staff will be housed. Significant work will go on at Wash U in the Lab. of Rob Mitra who was one of the early workers on the program.

INNOVATION: This rates very high on the scale of innovation. Access to single molecule data is a very powerful addition to the geneticists 'tool box' and with this in hand many methods can be addressed.

INVESTIGATORS: Mostly outstanding. The PI has a track record of amazing innovation but has not yet produced an enduring technology that is both revolutionary and enduring. Despite this he continues to innovate and this is in fact a strength of his participation in this program – he is not at all frightened of innovation. In this case there is every chance these polony structures will represent a 'home run' for him.

DATA RELEASE: This is a strength of the application. The real product of the application is the export of the technology. There is not a body of production data that will need to be disseminated. However the group is well known for software distribution and once built this will likely be performed properly.

SUMMARY: This is an exciting area of development and these methods and structures are likely to be used by others. Nevertheless there are many issues to be addressed before this is able to be used in the widest community. The group should make sure it focuses on the basics and get the underlying technology to be as robust and reproducible as possible. Outstanding success with one or a few applications where there are no competing methods would be a more impressive achievement than some demonstrations in areas where there are competing technologies.

The proposal aims to create the Molecular and Genomic Imaging Center. This is in response to the claimed community-wide need for flexible, cost effective high resolution technology to identify and characterize variation in biological systems.

CRITIQUE 2:

The background and significance sections provide compelling arguments for the development of inexpensive high throughput nucleic acids technology. The polony approach derives from Church's focus on miniaturization. The proposal makes the claim that the polony technology is realistic, close at hand, modular, and versatile. This proposal should be funded.

SIGNIFICANCE: The proposed research is generally highly significant, and could lead to several technological breakthroughs, including cost efficient whole genome sequencing and robust single cell nucleic acids technologies. If successfully implemented and supported, the technology and its uses would be extremely useful to many investigators, and would in my view expand the number of investigators who would use genomics to address their particular research interest.

APPROACH

Aim 1: The applicants have made progress in achieving the goals of their first aim. They have:

1. Developed chemistry to anchor one strand of every DNA molecule to the acryamide matrix.
2. Deposited up to 5 million distinguishable polonies on a single microscope slide.
3. Experimented with novel fluorescent nucleotides with cleavable linkers.
4. Achieved five to eight base reads on a small number of amplified polonies.

The objective of aim is to develop the capability for 20-500 base reads reliably from millions of polonies in parallel. Priorities for further development of the technology are to develop measures of sequence accuracy, improve sequence read length, undertake construction of template libraries and increase polony density.

The applicants indicate that have experience in construction of template libraries and that this technology is mature. No further discussion is given. I would have appreciated some discussion of the caveats of library construction with special consideration given to the polony technology.

The treatment of optimizing read length was not as strong as it could have been in my opinion. There was no clear discussion on what limits read length currently, and no explicit discussion on how to optimize read length beyond statements that alternate chemistries were to be explored. There was a passing comparison to pyrosequencing, a related sequencing-by-synthesis method. Polony sequencing is expected by the applicants to generate longer read lengths than pyrosequencing because there are no dilution-related issues and reaction can go to completion due to washing between cycles. Overall, there were clear implications that needed are new chemistries including new polymerases, but potential anticipated gains in read length from each of these areas was not discussed.

One factor presented as possibly limiting read length was cycle number, which in turn is limited by gel strength. The applicants dealt with this factor by using new polymers, commercial gel strengthening agents and gentler handling of gels. It was unclear to me what the impact on read length has been with these new approaches.

There was significant discussion on determining the number of incorporations at tandem mononucleotide repeats. This is generally well addressed, with potential problems and possible solutions presented.

Aim 2: Develop applications of polonies to gene expression profiling.

The condensed sequence tag (CST) approach has been implemented as an interim measure to applying polony technology to digital gene expression profiling. This is because currently the technology is unable to resolve homopolymers. The argument presented is that if CSTs are sufficiently long, an unambiguous match of the CST to the transcriptome or genome should be possible. A simulation is presented on p. 81, showing that CST lengths of 40 yields ~85% CSTs unique in genome.

The applicants claim this length is within reach as have generate 5bp reads with labeled nucleotides and 20 bp with unlabeled nucleotides. I feel their assertion is a realistic goal for the technology. However, the section on matching CSTs to transcripts was sketchy and without sufficient detail. It was unclear to me what the strategy for sequencing cDNAs was, and in particular the region of the cDNAs that will be sequenced. For example, will the applicants rely on a strategy comparable to that employed by SAGE? What then will the "matching" informatics strategy be? What databases of transcripts will be used to compare to CSTs? Will these be constructed in-house and if so what are the important considerations? What are the considerations in matching CSTs against the genome? How many CSTs are likely to be split by an intron-exon boundary? How will polymorphisms be handled? How will validation be accomplished?

Polony identification with pooled probes.

As an alternative to FISSEQ, the applicants propose to hybridize probe pools to polony slides to score for the presence of particular genes on the slide. Visualization can be accomplished by single base extension of fluorescently labeled nucleotides. 10 serial rounds of probe hybridizing are apparently possible, as is design of pooling strategies to facilitate multiplexing. For studies aiming to score for many genes, the reliance on lots of oligos may result in significant expense. Also, this approach scores only for the presence of genes you know about, which works against gene discovery property alluded to on page 81. Also unknown is robustness of pooling and ability to map individual polonies to transcripts - how do you know which probe went to which polony? It seems to me that this depends entirely upon ability to de-convolve probe pools. But what is the relationship between the number of pools you would need to perform vs. the number of genes you need to map to polonies? Can this not be estimated?

Scientific Application Neural differentiation of ES cells:

The applicants propose to make cDNA from 100 cells at each of 5 time points in ES cell differentiation, profile these using gene specific probes and then quantitate the entire transcriptome. This is a worthwhile scientific target, and could be used to demonstrate the efficacy of the approach, given that the results can be validated. What is the proposed mechanism by which the results will be validated? The aim to work towards single cell profiling, presumably for both sequencing and probing, is important.

Single molecule profiling of alt. pre-mRNA splicing:

In my view this is an important aim. The approach involves amplifying transcripts using "constant exons" as primer annealing sites, and then after denaturation, probing the transcript for specific exons.

An obvious problem is that one needs to know which are the constant exons. How will this be determined, and what are the consequences of failing to accurately predict this? The approach demands you have some knowledge of "candidate exons" ahead of time. Information on the order of exons in transcripts could conceivably be obtained, but the exact structure of each of the exons will not be known as these are not being sequenced. Hence, the hybridization approach has several limitations in this application. Another consideration concerns the mis-prediction of exons in current sequence annotations.

The proposal is to apply the technology to Tau and CD44 initially, and then a "larger set" of genes later on.

A primary advantage of the approach is that one is working from single molecules, not pools of molecules. Hence, rare spliced forms might not be diluted to the point where they are "swamped out" as they might be in applications sampling pools of molecules. A disadvantage of the approach is that it involves probing for known exons, resulting in a loss of discovery potential. Further, the process is serial. In my opinion, the ability to directly sequence transcripts would be much more appealing than probing transcripts for exon content, but probing is a reasonable intermediate aim. A challenge described in the proposal relates to scaling up the data acquisition process, both within and across labs.

Long Range Sequence Connectivity and Haplotyping:

For this aim, proof-in-principle has been established for genomic distances of 8 - 45 kb, and the applicants have published a PNAS paper on the topic. Key to this application is the identification of overlapping colonies, but how this is done and what the caveats are in doing it are not described well in the proposal, and also not well described in the PNAS paper. The dilution of template molecules seems key, with molecules dilute enough so that partially overlapping colonies can be confidently assigned to the same template, vs. random overlaps. Also it seems like the further targets are away on a chromosome, the easier it might be to identify overlapping colonies. There may be an optimal distance, where closer could negate effectiveness of overlapping colony identification.

The great strength of colonies in this regard seems to me is the ability to know phase. This will be demonstrated in the context of the on ApoE gene, which seems appropriate.

The applicants propose to extend this experimentation to long-range haplotyping. This will involve making 150 kb lengths of DNA, ligating these into BAC vectors, making colony slides and amplifying colonies with Phi29 (the claim is that this can amplify 150 kb DNAs). After amplification, a gel with millions of colonies will be produced, and both ends of each amplified BAC will be sequenced using FISSEQ. I wasn't clear on the method of immobilization of the single strands produced by Phi29 and the method of strand capture, but presumably this will be by the use of the modified primer. Also unclear is the effect that repetitive sequences will have on the effective number of useful FISSEQ sequences, but presumably it will reduce them by 30 - 50%.

A pooling scheme will be derived to facilitate typing of SNPs. The number of pooled primers is large. Questions that arose included: What is the expected primer concentration, and to what extent can these very large pools be constructed without interfering with hybridization efficiency. Another major acknowledged challenge is the repeated re-use of slides and gel stability.

High throughput de novo sequencing via colonies.

Sequencing of *H. pylori* is proposed, and the major challenge is sequencing of mate pairs, which will be addressed through strategies that don't depend on PCR. Strategies include use of jumping clones, multiple amplification from single molecules (MASMO) and Phi 29 amplification. These will be tested for amplicons longer than 10 kb. For amplicons less than 10 kb, rely on PCR but methods are not described. Instead, the applicants make reference to only a single paper without treating considerations.

Jumping clone libraries, or deleted insert libraries: Details are completely absent, but references to published protocols are given. The applicants take the stance that many groups have published on the

method and if they run into problems, they will consult with these authors. I found this a little too vague for my tastes.

MASMO: Technically challenging. The strategy makes sense however, and the approach is not limiting in sequencing the genome. A consideration is that it will be difficult to impose absolute size selection on PCR products, and so assembly approaches will not be able to rely heavily on prior knowledge of the sizes of template molecules.

Single cells and In situ profiling:

I found this section confusing and difficult to understand. Spatial resolution currently limited. For example, the applicants propose to "examine" (?) dsRNAs at single cell level. How is not at all clear - I felt some additional explanation would have been helpful. The applicants also propose to undertake multiplex multicellular profiling, looking for single stranded DNA and "priming elements" in cells undergoing chromosome co-segregation or random chromosome segregation. Here again I was confused. How will such cells be identified? What are the controls? What if such sequences are found? What if they are not? Is this likely to be a technological limitation or a real biological phenomenon?

Computational methods and Systems biology:

This section is absolutely key to the success of this CEGS, and involves several activities around developing software to enable generation and analysis of polony data. Activities include converting existing code to C++, improving base calling, achieving robotics integration, construction of LIMS and data management systems, derivation of sequence assembly algorithms, mathematical modeling of polony technologies, generation of error models, and undertaking systems approaches for integration of distinct data types. All are necessary and crucial, and here I became worried that insufficient resources had been allocated to this activity. There is enough to do that a small team (3 - 5?) of full time programmers could be kept busy!

MANAGEMENT: I saw no problems with the structure as proposed. Church oversees administrative and financial aspects of the proposed center. Adjustments to budgets will be made by co-directors, one of which is Church.

Progress will be monitored and guided at two levels.

(I) There will be monthly conference calls (or more frequent) involving "center contingents". PIs will meet more frequently as necessary to address issues. At least twice annually PIs will visit other sites (p. 123). An evaluation of progress towards the 5 aims will be made monthly by coPIs.

(II) Advisory board will be convened to review accomplishments and assess future directions. An annual report will be prepared by co-directors and distributed to the advisory board. Apparently lacking on the board is someone with software expertise, which is odd given the deep need for software development.

INNOVATION: The work is certainly innovative, and if nothing more than the sequencing eventually panned out a major contribution will have been made. The only difficulties I had were that in several parts of the proposal (as indicated above) some experiments were not explained to a satisfying level of detail, with the primary deficiencies related to the technical descriptions and level of detail offered in the description of the single cell / in situ work.

INVESTIGATORS: As is apparent in the Project Development Schedule, Church and Mitra are absolutely central to this effort. They invented polony technology and are imminently qualified and appropriate in their leadership roles on this proposal. Gottlieb and Sherley provide the specialized expertise around the specialized biological applications, such as stem cell analysis and alternate splicing. They are excellent scientists, but in my opinion the main thrust of this is the development of technology and not so much the application of early forms of the technology to biological problems. There will be many applications to biology if the technology evolves.

ENVIRONMENT : The environment for the proposed research is excellent.

DATA RELEASE: Plans as proposed are adequate, and Church has a record in provision of resources. Data and computer programs will be made available from the MGIC website. Documented source code will be copyright and made available freely to academics and licensed for commercial use. A rapid patenting and publishing strategy is described for moving discoveries into the public domain.

TRAINING: The training component is well described and strong. The applicants propose to create training opportunities by sponsoring a center-specific course, a center-based seminar series and by supporting bridging research positions that will allow trainees to work closely with multiple center investigators. They will also sponsor mini-courses to enable communication to scientists outside the specialty of the investigators. Further, they propose to host training seminars on polony - related protocols. Video conferencing will allow participation in seminar series at both Boston and St. Louis sites.

Five minority undergraduate students will be actively recruited for annual summer training opportunities in each of the laboratories, with 2 students in the Church lab. This will be accomplished by working with existing programs at the Universities. The applicants also propose to mentor certain of the above students working on other campuses after the summer studentship. A specific example of such a relationship involving Church is given. Finally the applicants propose to establish one month mini sabbaticals to bring in faculty and other researchers from minority serving institutions to stimulate new research / funding opportunities.

CRITIQUE 3:

The application is for a P50 Center of Excellence in Genome Science (CEGS) grant to fund a research project "Molecular and Genomic Imaging Center". Representatives from 4 institutions serve as PI and co-investigators on this project, including Dr George Church of Harvard Medical School listed as PI. The CVs submitted show these individuals include well-established researchers in genome technologies, computational biology, and molecular biology capable of executing the challenging project described in the application.

The CEGS application proposes to expand on the nascent polony technology developed in the Church lab, to increase the utility and robustness of the polony assay through improved chemistry and analysis software. The polony technology described in its present form generates a "PCR colony" in a thin acrylamide gel, where one strand of the PCR product is covalently bound to the acrylamide matrix. The resulting polonies occur at high density, potentially millions on a slide, and can be probed with fluorescent oligos or sequenced by *in situ* sequencing methodology. As this innovative platform is developed, it may enable extremely high throughput sequence detection, genotyping, and sequence tagging at far lower cost than possible now. The applicants envision the development of a robust platform that can be disseminated to a wide range of users at low cost for use in a large variety of applications requiring sequencing, SNP detection, haplotype analysis, mRNA profiling, and detection of sequences in single cells or small tissue regions. The technology right now appears to have basic capabilities, such as the ability to detect exons by oligo hybridization, the ability to generate short (16

base) sequence reads, and the capability to identify SNPs separated by 10s of kb intervening sequence for long-range haplotype analysis. In its present form, the technology has been used in a handful of publications and appears on the verge of enabling significant, perhaps breakthrough, contributions to areas of research requiring genome scale technologies, such as: genetic association studies, identification of rare mutations within a population, exploration of splice variants, and an enhanced SAGE approach to analysis of mRNA transcript abundance. By supporting this CEGS we can be more assured that the technology will become more robust, more broadly distributed, and I suspect more applications will be discovered.

The proposal also endeavors to enhance the technology, by increasing the *in situ* sequencing read lengths, attempting multiplex and pooling strategies for sequence and SNP detection, increasing the length of target sequences bound to the matrix, and generation of colonies from single cells. While the goals are ambitious, incremental improvements toward each of these technical goals seem plausible and will enable a larger range of applications. An example of a challenge in the application is to increase the length of *in situ* sequence reads to around 100 bases per primer, and to improve the ability to measure and discriminate single base runs. Any significant application of the polony technology in resequencing will probably require this read length enhancement (and even more I suspect), however, any improvement will have a significant impact on the type of projects where this technology can be employed. One of the aspects I like about this proposal is that even if the challenging goals are not fully met, incremental technology enhancements will be of high value and utility to a growing community of users. This is an exciting technology, perhaps on par with the development of PCR. I want to try it out myself, using published protocols, as soon as I can!

The proposal has a good plan for governance and administration, and plans to share data, results, software tools, and training through a web site. A web site exists now to share polony technology with a wider audience.

THE FOLLOWING RESUME SECTIONS WERE PREPARED BY THE SCIENTIFIC REVIEW ADMINISTRATOR TO SUMMARIZE THE OUTCOME OF DISCUSSIONS OF THE REVIEW COMMITTEE ON THE FOLLOWING ISSUES:

PROTECTION OF HUMAN SUBJECTS (Resume): ACCEPTABLE The project qualifies for Exemption Category 4, as the research involves using existing data or specimens that have no links to the original subjects.

INCLUSION OF MINORITIES PLAN (Resume): ACCEPTABLE - There is no information regarding race or ethnicity linked with the existing data or specimens used in the study and that is acceptable.

INCLUSION OF WOMEN PLAN (Resume): ACCEPTABLE There is no minority information regarding gender linked with the existing data or specimens used in the study and that is acceptable.

INCLUSION OF CHILDREN PLAN (Resume): ACCEPTABLE There is no information regarding age connected with the existing data or specimens used in the study and that is acceptable.

VERTEBRATE ANIMAL (Resume): Methods for use of vertebrate animals are adequately described and the committee had no concerns.

COMMITTEE BUDGET RECOMMENDATIONS: The budget was recommended as requested.

MINORITY ACTION PLAN ASSESSMENT (RESUME): Overall, the Minority Action Plan was considered quite good, taking advantage of ongoing programs and infrastructure at Harvard Medical School, MIT, and Washington University. At Harvard, the program is strengthened by the PI's

connection with such groups as the Society for Advancement of Chicanos and Native Americans in Science (SACNAS) where he serves on the Advisory Committee, but no information was provided about the other investigator's previous activities. A general weakness of the plan is that while it describes the activities at Harvard, relatively little information is provided about specific programs at the other two sites. The same seemed true about the investigators. The commitment of Dr. Church to these activities seemed quite clear, but no information was provided about the other investigators.

Four types of activities are proposed: a summer undergraduate experience, collaborative research efforts, post-college opportunities, and mini-sabbaticals for faculty at minority institutions. The Center will tie into ongoing undergraduate summer research experiences at Harvard, MIT, and Washington University and proposes each year to recruit 5 students into the laboratories of 4 key investigators. The collaborative research effort is aimed at continued support for summer research projects of promising students that would be carried out at the student's home institutions. For example, funds are requested for a student at the University of Puerto Rico who began a summer research project in Dr. Church's laboratory. Similar longer-term mentoring relationships are proposed for the other key investigators. The post-college opportunities will support recent graduates and alumni of summer research programs to pursue one or two of years of research experience and GRE preparation courses before applying to graduate school. Finally a mini-sabbatical program will support a one-month sabbatical to bring faculty from a minority serving institution to one of the Center laboratories. This component was not well developed and it seemed to the committee that the one-month period was too short and the level of support too low.

To manage these activities, the Center proposes a Director of Minority Programs, and a half-time Coordinator of Minority Initiatives. For the Director, the Center has recruited Dr. Jocelyn Spragg who has considerable experience with programs promoting women and minority participation in the sciences. It was unclear if these individuals will have a role in managing the activities at MIT and Washington University activities. If so, how the management structure will be set up? If not, who will manage the efforts at these sites? Given all the activities, it seemed that a full-time coordinator is warranted.

ASSESSMENT #1

1. Summary of Action Plan Response

- Are the goals of the program clear and will the objectives and activities as outlined accomplish the goals? --This minority action plan intends to increase the number of minorities involved in genomic research programs through a Molecular and Genomic Imaging Center (MGIC). They will build upon existing infrastructures associated with Harvard Medical School, MIT, and Washington University, they will recruit individuals that might not normally consider our areas of endeavor and, in this way, broaden the intellectual strength of genomic science. Four types of training are envisioned, a summer undergraduate experience, collaborative research efforts, which will deepen the summer experience for some students, post-college opportunities, and mini-sabbaticals for faculty at minority institutions. This is a reasonably good plan, with some detail.
- How well is genomics integrated into planned activities? -- Genomics will be used as research expertise to train the in all activities described.
- If PI plans to participate in an ongoing activity designed and managed by others, what is the "value added" of their participation? -- The PI will participate as a trainer in all proposed activities. A Director of Minority Initiatives at Harvard has been identified who already works in these programs, Dr. Jocelyn Spragg. Fifteen percent of her time is requested; and in addition a minority coordinator will be hired in a half time position.

Activity 3: Post-College Opportunities

- What are the strengths and weaknesses of the activity? -- The strength of this activity is that it follows a established pattern observed at Harvard, that some summer research participants come back to work in the labs before joining graduate school. The investigator is attempting to take advantage of this fact and offer incentives for the students to apply to graduate school such as tuition free attendance of courses, and a paid GRE training activity. The observed weakness is that we do not know if they will be able to recruit the six desired individuals.
- How could the activity be improved? What elements should be included in this activity to make it an effective program? -- Formalizing this activity as a Post –Bac program would give more legitimacy and individuals could be recruited into the program.
- Will the activity facilitate participants moving to the next phase of their educational or career program and if so, how? -- Yes, directly via GRE prep courses; indirectly via contacts, which will be helpful to the student as he/she applies to graduate school.
- Are the milestones appropriate? If not, how can they be refined? -- It is uncertain whether the six students can be recruited.
- Is the evaluation component appropriate? If not, how can they be refined? -- No evaluation described
- Are individuals with the right expertise involved in the development and management of the program? -- Yes
- Will this activity facilitate the long-term goal of the NHGRI Action Plan? -- Yes
- Is the budget appropriate? -- Yes

Activity Rating: ___Acceptable X Acceptable With Modifications ___Revise
 ___Should Not Be Pursued

Activity 4: Mini-sabbaticals

- What are the strengths and weaknesses of the activity? -- Excellent activity, which again builds on the connections of the PI and an HBCU, Morehouse School of Medicine. The activity is strong and will provide the desired outcome. Weaknesses reflect the lack of involvement of the other institutions.
- How could the activity be improved? What elements should be included in this activity to make it an effective program? -- The activity is not defined, but it will improve as connections between Harvard and Morehouse are strengthened.
- Will the activity facilitate participants moving to the next phase of their educational or career program and if so, how? -- It will facilitate the acquisition of good skills by faculty members in this area as well as exposure to a rich research environment.
- Are the milestones appropriate? If not, how can they be refined? -- Two faculty are appropriate to start the activity.
- Is the evaluation component appropriate? If not, how can they be refined? -- No evaluation
- Are individuals with the right expertise involved in the development and management of the

program? -- Yes

- Will this activity facilitate the long-term goal of the NHGRI Action Plan? -- Yes
- Is the budget appropriate? -- No. Sabbatical costs will be much higher than \$5,000/year and travel. Also it may take more than one month to complete the activity.

Activity Rating: ___Acceptable X Acceptable With Modifications ___Revise
 ___Should Not Be Pursued

Assessment #2

1. Summary of Action Plan Response

- Are the goals of the program clear and will the objectives and activities as outlined accomplish the goals? -- There are no specific aims for this component in the abstract. Training and minority inclusion are written together. Will do seminars and workshops and those that are successful will be opened to the general community. They should give some sense of what their previous success has been in terms of training minority students.
- How well is genomics integrated into planned activities? -- Fairly well. Labs will have one summer minority student each, but there should be some effort on getting minority graduate students and post docs?
- How does the plan take advantage of the research infrastructure? -- There will be Center-related courses and seminar series in Boston and at Washington University. The advisory committee needs to have more people who have practical experience working with these types of program such as a SACNAS faculty member.
- Is the level of funding for the proposed training activities commensurate with the requested level of funding for the entire project? -- The applicants say it is 10%, but there is no breakdown of costs. The one-month sabbatical at \$5,000 translates to a 9-month salary of \$45,000. Also, I would hope they would open up these interactions to involve faculty from other schools. I think they need a full time coordinator who can help assess and change or bring in new program initiatives. I wish there would be a component that would allow Harvard and Wash U faculty to get to know more about minority cultures, e.g., spend a week in a pueblo or on a reservation.

Summary: More information is needed about the number of students mentored in each laboratory and how successful the labs have been in mentoring minority graduate students or post docs. There is a need for a full time coordinator. Harvard remains a difficult place for minority students to gain access, and there needs to be some creative ways to induce change in the environment.

OVERALL RATING: ___ ACCEPTABLE ___ ACCEPTABLE WITH MODIFICATIONS
 X NEEDS SIGNIFICANT REVISION

2. Assessment of Specific Activities

Activity 1: Summer Undergraduate Experiences

- What are the strengths and weaknesses of the activity? -- This is an established, successful program. However, what does "usually positive" evaluations mean in percentages. What changes have been made as a result of the evaluation?
- How could the activity be improved? What elements should be included in this activity to make it an effective program? -- Minority programs, especially for Native Americans, seem to lack administrative support. Specific details about how such minority students will be recruited are

needed.

- Will the activity facilitate participants moving to the next phase of their educational or career program and if so, how? -- There isn't any planned mentoring activity. It is expected that this will happen, but the students who will work with the minority students need to be part of this, as well.
- Are the milestones appropriate? If not, how can they be refined? -- Yes, but there is a need to think about placement of cohorts so the students don't feel isolated.
- Is the evaluation component appropriate? If not, how can they be refined? -- There is no evaluation component.
- Are individuals with the right expertise involved in the development and management of the program? -- Probably, but it would be useful to get more details about what evaluation has taken place, what are common criticisms of the program, and what measures have been taken to deal with concerns.
- Will this activity facilitate the long-term goal of the NHGRI Action Plan? -- The presentation could have been stronger.
- Is the budget appropriate? -- No, it doesn't include funds or travel and there is no money for recruitment. As a result, only people going to a few meetings will hear about the program.

Activity Rating: ___Acceptable Acceptable With Modifications ___Should Not Be Pursued

Activity 2: Collaborative Research Efforts

- What are the strengths and weaknesses of the activity? -- There is a faculty person from Puerto Rico who has developed a collaboration with the Church lab. Other faculty will develop mentoring relationships by going to meetings. It is not clear how student will be selected.
- How could the activity be improved? What elements should be included in this activity to make it an effective program? -- This is a nice idea. I think it would be good for the faculty to identify minority faculty in different parts of the country for this relationship – not all in Puerto Rico.

Activity Rating: ___Acceptable Acceptable With Modifications ___Should Not Be Pursued

Activity 3: Post-College Opportunities

- What are the strengths and weaknesses of the activity? -- This program will pay for the students to attend a class and take the GREs, but it is not specified who will pay them to do their research. It is assumed that each lab will cover that expense.
- Is the budget appropriate? Unclear, as it depends on whether there is funding from other sources for the students to do research.

Activity Rating: ___Acceptable ___Acceptable With Modifications ___Should Not Be Pursued

Activity 4: Mini-sabbaticals

- What are the strengths and weaknesses of the activity? -- Not enough money per faculty member.
- How could the activity be improved? What elements should be included in this activity to make it an effective program? -- It is unclear what the mentoring situation will be. Can they come from institutions other than Morehouse?

- Are the milestones appropriate? If not, how can they be refined? -- Two faculty per year.
- Is the budget appropriate? No, not enough money for each sabbatical.

Activity Rating: ___Acceptable XAcceptable With Modifications ___Should Not Be Pursued

KDN

NOTICE: The NIH has modified its policy regarding the receipt of amended applications. Detailed information can be found by accessing the following URL address:
<http://grants.nih.gov/grants/policy/amendedapps.htm>

NIH announced implementation of Modular Research Grants in the December 18, 1998 issue of the NIH Guide to Grants and Contracts. The main feature of this concept is that grant applications (R01, R03, R21, R15) will request direct costs in \$25,000 modules, without budget detail for individual categories. Further information can be obtained from the Modular Grants Web site at <http://grants.nih.gov/grants/funding/modular/modular.htm>

MEETING ROSTER

**Genome Research Review Committee
National Human Genome Research Institute Initial Review Group
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
GNOM-G 1
November 17, 2003**

CHAIRPERSON

GIBBS, RICHARD A., PHD
DIRECTOR AND WOFFORD CAIN PROFESSOR
HUMAN GENOME SEQUENCING CENTER
DEPARTMENT OF MOLECULAR AND HUMAN GENETICS
BAYLOR COLLEGE OF MEDICINE
HOUSTON, TX 77030

MEMBERS

CHERRY, J. M, PHD *
ASSOCIATE PROFESSOR (RESEARCH) IN GENETICS
DEPARTMENT OF GENETICS
STANFORD UNIVERSITY SCHOOL OF MEDICINE
STANFORD, CA 943055120

ENDY, DREW , PHD *
FELLOW
DEPARTMENT OF BIOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA 02139

JIRTLE, RANDY L, PHD *
PROFESSOR
DEPARTMENTS OF RADIATION ONCOLOGY &
PATHOLOGY
DUKE UNIVERSITY
DURHAM, NC 27710

KARP, PETER D., PHD
DIRECTOR, BIOINFORMATICS GROUP
SRI INTERNATIONAL
MENLO PARK, CA 94025

KARP, RICHARD M, PHD
UNIVERSITY PROFESSOR
DEPARTMENTS OF BIOENGINEERING, AND ELECTRICAL
ENGINEERING & COMPUTER SCIENCE AND
MATHEMATICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720

LIMA, MARIA F., PHD
PROFESSOR
DEPARTMENT OF MICROBIOLOGY
DEAN, SCHOOL OF GRADUATE STUDIES AND RESEARCH
MEHARRY MEDICAL COLLEGE
NASHVILLE, TN 37208

MARRA, MARCO A., PHD *
DIRECTOR
GENOME SCIENCES CENTRE
BC CANCER AGENCY
VANCOUVER, BC V5Z 4E6
CANADA

MORROW, BERNICE E., PHD
ASSOCIATE PROFESSOR OF MOLECULAR GENETICS
DEPARTMENT OF MOLECULAR GENETICS
ALBERT EINSTEIN COLLEGE OF MEDICINE
NEW YORK, NY 10461

NEUBAUER, MICHAEL G., BA
GROUP LEADER
APPLIED GENOMICS DEPARTMENT
BRISTOL-MYERS SQUIBB PHARMACEUTICAL
RESEARCH INSTITUTE
PENNINGTON, NJ 08543

NICHOLLS, ROBERT D., DPH, DPHI, PHD *
PROFESSOR OF GENETICS IN PSYCHIATRY
CENTER FOR NEUROBIOLOGY AND BEHAVIOR
DEPARTMENT OF PSYCHIATRY
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA 191046140

SZALLASI, ZOLTAN , MD *
ASSISTANT PROFESSOR
CHILDREN'S HOSPITAL INFORMATICS PROGRAM
HARVARD MEDICAL SCHOOL
BOSTON, MA 02115

WERNER-WASHBURNE, MARGARET C., PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF BIOLOGY
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NM 87131

SCIENTIFIC REVIEW ADMINISTRATOR

NAKAMURA, KEN D., PHD
SCIENTIFIC REVIEW ADMINISTRATOR
SCIENTIFIC REVIEW BRANCH
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

* Temporary Member. For grant applications, temporary members may participate in the entire meeting or may review only selected applications as needed.

Consultants are required to absent themselves from the room during the review of any application if their presence would constitute or appear to constitute a conflict of interest.