NATIONAL ADVISORY COUNCIL FOR HUMAN GENOME RESEARCH
SUMMARY OF MEETING[1]
February 11, 2008

The open session of the National Advisory Council for Human Genome Research was convened for its fifty-second meeting at 8:35 A.M. on February 11, 2008 at the Fishers Lane Conference Center**,** Rockville, MD. Francis Collins, Director of the National Human Genome Research Institute, called the meeting to order.

The meeting was open to the public from 8:35 A.M. until 3:15 P.M. on February 11, 2008. In accordance with the provisions of Public Law 92-463, the meeting was closed to the public from 3:15 P.M. on February 11, 2008 until adjournment for the review, discussion, and evaluation of grant applications.

Council members present:

Eric Boerwinkle
Andrew Clark
Jorge Contreras, Jr.
Sean Eddy
Vanessa Northington Gamble
Richard Gibbs
Geoffrey Ginsberg
Caryn Lerman
Deirdre Meldrum
Patrice Milos
Jeffrey Murray (participating via teleconference)
David Page
Stephen Prescott
Paul Sternberg
David Valle
Richard Weinshilboum

Ex Officio member absent:

Gerard Schellenberg

[1] For the record, it is noted that to avoid a conflict of interest, Council members absent themselves from the meeting when the Council discusses applications from their respective institutions or in which a conflict of interest may occur. Members are asked to sign a statement to this effect. This does not apply to "en bloc".

Staff from the National Human Genome Research Institute:

Ajay, DER
Glory Baldwin, DER
Tsegahiwot Belachew, DER
Catherine Bennet, DER
Les Biesecker, DIR
Vivien Bonazzi, DER
Vence Bonham, OD
Joy Boyer, DER
Lisa Brooks, DER
Comfort Browne, DER
Debbie Chen, DER
Cheryl Chick, DER
Monika Christman, DER
Francis Collins, Director
Camilla Day, DER
Karen DeLeon, OD
Laura Dillon, DER
Gwen Dudley, DEAS
Carla Easter, OD
Elise Feingold, DER
Adam Felsenfeld, DER
Kimberly Ferguson, DER
Colin Fletcher, DER
Phyllis Frosst, OD
Barbara Fuller, OD
Peter Good, DER
Eric Green, DIR
Bettie Graham, DER
Alan Guttmacher, OD
Mark Guyer, DER
Emily Harris, OD
Lucia Hindorff, OD
MK Holohan, OD
Diane Johnson, OD
Chris Juenger, DER
Heather Junkins, OD
Mike Kabatt, DER
Kathy Kopnisky, OD
Carson Loomis, DER
Teri Manolio, OD
Jean McEwen, DER
Keith McKenney, DER
Lisa McNeil, OD
Anika Mirick, DER
Janis Mullaney, OD
Ken Nakamura, DER
Brad Ozenberger, DER
Carmen Perera, DEAS
Jane Peterson, DER
Anne Pierson, DER
Rudy Pozzatti, DER
Tammy Rabinovich, OD
Erin Ramos, OD
Eddie Rivera, OD
Jerry Roberts, DER
Cristen Robinson, DER
Laura Rodriguez, OD
Anna Rossoshek, DER
Jeff Schloss, DER
Geoff Spencer, OD
Jeff Struewing, OD
Carolyn Taylor, DEAS
Gary Temple, DER
Elizabeth Thomson, DER
Larry Thompson, OD
Jasse Tucker, OD
Susan Vasquez, OD
Lu Wang, DER
Vivian Ota Wang, DER
Chris Wellington, DER
Kris Wetterstrand, DER
Diane Williams-Bey, DEAS

Others present for all or a portion of the meeting:

David Altshuler, Broad Institute
Diane Baker, Genetic Alliance
Joann Boughman, American Society of Human Genetics
Bill Gelbart, Harvard University
Daniela Gerhard, NCI
Richard Gibbs, Baylor College of Medicine
Rodney Howell, American College of Medical Genetics
Eric Lander, Broad Institute
Rick Lifton, Yale University
Sharon Olsen, International Society of Nurses in Genetics
Sharon Terry, Genetic Alliance
Wendy Uhlmann, National Society of Genetic Counselors
George Weinstock, Washington University
Alan Williamson, Consultant
Rick Wilson, Washington University
Barbara Wold, California Institute of Technology

INTRODUCTION OF NEW MEMBERS AND STAFF, LIAISONS AND GUESTS

The major agenda item for the Open Session was a discussion of Staff proposal for reprioritization of the large-scale sequencing program in response to new opportunities to utilize next generation sequencing technologies. A large number of guests were invited to attend this meeting to participate in this discussion.

Dr. Guyer introduced several *ad hoc* Council members:

Geoffrey Duyk, former Council member and member of the Sequencing Advisory Panel (SAP); Sean Eddy, former Council member; Geoffrey Ginsberg, newly appointed Council member; Jeff Murray, former Council member; and David Valle, newly appointed Council member.

Pilar Ossorio, another newly appointed Council member, was unable to attend the meeting.

Dr. Guyer also introduced several guests:

Bill Gelbart, Rick Lifton, Alan Williamson, and Barbara Wold are all SAP members, as well as former Council members; David Altshuler and Les Biesecker, co-chairs of the Medical Sequencing Working Group; Eric Green, NISC Director; Eric Lander, Broad Institute Director; George Weinstock, formerly the co-Director of the Baylor Human Genome Sequencing Center, currently at the Washington University Genome Sequencing Center (GSC), and also a former Council member; and Rick Wilson, Director of the Washington University GSC.

Dr. Guyer introduced new NHGRI staff:

Tsegahiwot Belachew, Scientific Program Analyst; Priscilla Crocket, Administrative Officer; Camilla Day, Scientific Review Administrator; Kimberly Ferguson, Grants Financial Analyst; Zephaun Harvey, Grants Management Specialist; Lucia Hindorff, Program Director; and Heather Junkins, Scientific Program Analyst.

Dr. Francis Collins introduced several new staff members of the Office of the Director:

Diane Johnson, Program Assistant; Kathy Kopnisky, Chief of Staff; June Langley, Executive Assistant to the Director; Janis Mullaney, Executive Officer; and Tammy Rabinovich, Administrative Assistant.

Dr. Guyer welcomed members of the press and liaisons from professional societies:

Diane Baker, Genetic Alliance; Rodney Howell, American College of Medical Genetics; Sharon Olsen, International Society of Nurses in Genetics; Sharon Terry, Genetic Alliance; and Wendy Uhlmann, National Society of Genetic Counselors.

Dr. Guyer also welcomed Daniela Gerhard from NCI.

APPROVAL OF MINUTES

The minutes from the September 2007 Council meeting were approved as submitted.

FUTURE MEETING DATES

The following dates were proposed for future meetings: May 19-20, 2008, September 8-9, 2008, February 9-10, 2009, May 18-19, 2009, September 14-15, 2009, and February 8-9, 2010.

REPRIORITIZING THE LARGE-SCALE SEQUENCING PROGRAM

Introduction

Dr. Collins noted that, since the sequencing discussion was the major agenda item for discussion in the Open Session and likely to occupy the entire morning, we would proceed directly into that discussion and postpone the Director's Report until after lunch. He then introduced the discussion by reporting that, at its December meeting in St. Louis, the Sequencing Research Network had discussed an issue raised by NHGRI staff about the value of reprioritizating of the large-scale sequencing program to address the effective introduction of the new, disruptive sequencing technologies. The PIs were supportive of the idea and were then asked to prepare proposals for how they would revise their programs to use up to 50% of their awards over the next twelve months to focus on those projects that would be most well-suited to driving the implementation of the new technologies. Those proposals were received by staff and discussed with the Sequencing Advisory Panel, and the overall plan is now being brought to Council for consideration.

Background

Dr. Adam Felsenfeld gave an overview of the current state of the large-scale sequencing program. The majority of production capacity consists of ABI 3730 instruments that are primarily used in non-directed sequencing strategies; small amounts of the ABI 3730 capacity is used for directed sequencing projects, and some 454 sequencing is also done. Many of the projects for sequencing the genomes of additional organisms that are currently on the priority list are planned for ABI 3730 non-directed sequencing. However, many of the high priority projects that are currently in the planning or early implementation stages, including medical and human sequencing, will either need to utilize or benefit from utilizing the new sequencing technologies. Dr. Felsenfeld also provided a response to a request from the September 2007 Council meeting for an assessment of the impact of the large-scale sequencing program on biomedical research. He presented data on the number of journal citations of the several large-scale sequencing papers, as well as the number of unique journals in which the sequencing papers were cited, the number of patent citations and the number of database hits for genomes assembled and released through the program. He concluded that, while these metrics do not fully evaluate the program, they show a trend of scientific utility for the large-scale sequencing program projects.

High Priority Projects

Cancer Genome Sequencing Projects (TSP and TCGA) – Dr. Rick Wilson

The Tumor Sequencing Project (TSP) is an NHGRI sequencing project that began in 2005 as a developmental bed for scaling up PCR-based, ABI3730 sequencing. It is a collaboration among the three NHGRI large-scale sequencing centers and four cancer centers. Part A involved sequencing 600 genes of interest (exons only) in about 200 lung adenocarcinoma samples. The Consortium's initial SNP array analysis of the tumors identified a number of new sequencing targets in regions of amplification and deletion (these results were reported a paper published in *Nature* in December 2007). Further analysis is currently under way, and preliminary results suggest identification of mutations in four genes that show a positive correlation with tumor stage, mutations in 4 other genes correlate with sub-type, and mutations in other genes correlate with smoking history.

For TSP Part B, each sequencing center has a different plan for integration of the new sequencing technologies into the sequencing of tumor genomes: the Washington University Genome Sequencing Center plans to use the Solexa technology and 454-ready cDNA libraries to look at alignment, coverage, point mutation, indel detection and alternative splicing for transcriptome analysis; the Baylor Human Genome Sequencing Center plans to take an exon sequencing approach using Nimblegen capture technology and 454 sequencing; and the Broad Institute plans use a solution hybridization selection to capture the top 75 mutated genes from the initial TSP study and an additional 400 exons with poor ABI coverage, followed by Illumina sequencing, in 42 tumor/normal pairs chosen from across the stromal contamination bins.

The Cancer Genome Atlas (TCGA) is a collaborative project between NHGRI and NCI that is analyzing three different tumors in a pilot project. TCGA involves the analysis of DNA and RNA extracted from qualified tumor samples at a Biospecimen Resource Center by three NHGRI-sponsored large-scale sequencing centers and seven NCI-supported Cancer Genome Characterization Centers. The first cancer to be studied is glioblastoma multiforme. The sequencing centers began by sequencing 601 candidate genes in 96 glioblastoma samples; the genes were chosen by committee of GBM experts to address hypotheses based on the literature and glioblastoma biology. A second set of 725 sequencing targets has now been chosen on the basis of the initial genome characterization data from the Cancer Genome Characterization Centers (based on an analysis of copy number and expression), conserved regions across evolution, and tumor-specific splice variants.

In response to specific questions from Council members about TCGA, Dr. Wilson made the following additional points:

- There has been no validation of the sequencing results with clinical data yet.
- The different centers are using different approaches to mutation validation.
- There should be enough DNA for whole-genome sequencing in the case of some of the GBM samples and a pilot will be staring on that soon.
- New approaches to investigate small, non-coding RNAs are being evaluated.
- The exon capture methodologies being piloted will not pick up alternative splice sites and, therefore, it is important to use multiple approaches.
- Consent for the transcriptome studies has been obtained in the case of Dr. Wilson's AML work; for GBM, living patients will be reconsented for all types of genomic analysis, while samples from deceased individuals will only be used where the consent was acceptable for such studies.
- There are not yet enough data to allow analyses of cases in which there are mutations in more than one gene.
- There is no indication that RNA integrity will be a problem for the transcriptome studies.
- There are still many challenges in implementing the new sequencing technologies, but at the moment, things are going well with the Roche/454 and Illumina/Solex instruments but implementation of the ABI SOLiD instrument (which hasn't been available as long as the others) has been more difficult.
- It is too early to draw any conclusions about the relationship of differences in EGFR mutations to differences in tumor biology.

Human Microbiome Project (HMP)

Dr. George Weinstock presented an overview of the HMP, which is an NIH Roadmap initiative that is being co-led by NHGRI, along with NIAID, NIDCR, and NIDDK. He began by noting that this project would not be possible without the new sequencing technologies. Every human is colonized by thousands of species of microorganisms, and the objective of the HMP is to determine how to analyze and study the microbial content of human body sites and how to relate that to health and disease. Three types of projects will be involved in the HMP: sequencing of genomes of isolated organisms as reference

for other studies, development of a database of 16S sequences, and metagenomic analysis of microbial populations. In addition to the sequencing centers at Baylor, Broad and Washington University, the J. Craig Venter Institute's sequencing program is also participating in the HMP.

The overall experimental design of the HMP is complex and involves many components. Through its large-scale sequencing program, NHGRI has supported some limited initial efforts to sequence some references strains and metagenomic samples. The Roadmap has provided some "jumpstart" funds to sequence additional reference genomes and to recruit an initial cohort of individuals for sampling and metagenomic sampling of several body sites using 16S ribosomal sequencing. The overall HMP program will also include a set of demonstration projects that will investigate the linkage between the microbiome and health and disease, and a technology development component.

To date, about 100 reference genomes have already been sequenced; the goal is to complete (worldwide) a total of 1000 reference sequences as high-quality drafts and to finish 15% of them. The current costs for generating a high-quality draft of a bacterial genome sequence are estimated to be about $25,000 by Sanger sequencing and $20,000 by 454-based sequencing. Future reductions in the cost of 454 sequencing (from a combination of many factors, including reduction in the coverage necessary, more bases per run, and reduced reagent costs) could potentially drive the cost down to hundreds of dollars per genome. At that point, other bottlenecks, such as strain acquisition, would become more of a problem than sequencing cost.

In response to questions from Council members, Dr. Weinstock made the following additional points:

- Metadata will be associated with the research samples.
- Normalization strategies are being included in the technology development component of the HMP. Capturing the relative stoichiometry of bacteria at different body sites is not an articulated goal of the HMP but should be included in follow-up studies.
- Capture strategies have not been used to probe deeply into 16S, but their chips have been developed for this purpose.
- Samples of host DNA are being collected for future analysis.
- Metatranscriptomics is another area in which technology development projects would be appropriate.
- Model organism studies are not included in the HMP at this time, but there is a possibility that such studies could be funded as demonstration projects.

1000 Genomes

Dr. David Altshuler presented a general overview of medical sequencing projects with a specific focus on the 1000 genomes project.

One of the major hypotheses that has governed NHGRI funding is that systematic studies of genomic sequence variation may be used to find the genetic contributions to disease.

There have been many examples of success in associating variation in specific loci with common disease, but the actual genes have been identified based on the association data in only a few cases; the underlying basis of the vast majority of heritability of common disease is still not known. Thus, the challenges remain to find all of the rare and common causal mutations in loci identified in genome-wide association studies, and to develop methods to find genomic loci that play an important role in disease, but for which no common variants have been identified. NHGRI has taken two approaches to address the full range of mutations that cause disease.

The Medical Sequencing Working Group has provided advice to the Council and the Institute on sequencing in phenotyped samples, and has focused on rare diseases that haven't been solved by standard approaches and on common diseases through the determination of the full allelic spectrum at causal loci.

The 1000 Genomes Project is designed to extend the database of genetic variants that is already available, to include all SNPs present at a frequency of ≥1% or greater across the genome and at ≥0.1% in exons, as well as structural variation, and to generate accurate genotype and LD information to allow imputation. The project will probably utilize a diversity of next generation platforms. Challenges include: the use of diverse samples, optimization of technology platforms and data management and analysis. A number of working groups have been established within the project to address these issues. The project has begun as a pilot to generate enough data as the basis for determining a statistically robust strategy for the project.

In response to questions from Council members, Dr. Altshuler made the following additional points.

- To address the ethical challenges that the project is facing, the 1000 Genomes Project will initially use the HapMap samples that are properly consented for sequencing and release of data. The Project has also discussed the difference between the release of SNP data and full sequence data, and the release of phenotype data.
- To address the balance between data production and technology development, the participation of both the data producers and analysts as co-equals in determining project direction allows the identification of issues and the ability to make mid-course corrections.
- Acknowledging that there are many misconceptions in the community about the project and release of data, all data will be released and the samples from HapMap are already available at the Coriell Institute. In this regard, there has been confusion between the 1000 Genomes Project and a Chinese project that is also surveying genomic variation in Chinese populations.

Opportunities for New Technology Implementation

Genome-wide resequencing

Dr. Rick Wilson presented an overview of genome-wide resequencing and the next generation sequencing technology implementation, including a rough comparison of current production costs for the ABI 3730 (6-fold coverage needed for a draft genome), Roche/454 FLX (12-fold needed), and Illumina/Solexa (25-fold needed). The cost of a 3000 Mb genome would be around $15,000,000, $2,500,000 and $550,000, respectively, and ten-fold lower for a 300 Mb genome. Dramatic cost savings have already been achieved in the sequencing of the *Biomphalaria glabrata* genome; with the use of a mixed strategy including Roche/454 reads, the cost dropped from $4.29M to $2.50M. The new sequencing technologies can also be implemented in mutation detection (Roche/454 FLX) and massively parallel whole genome sequencing (Illumina/Solexa). The large-scale sequencing centers agree that a mixture of the different sequencing technologies will be implemented for different sequencing projects. Dr. Wilson's summary conclusions were:

- "Next gen" sequencing is here and will have a substantial effect on the sequencing of cancer genomes.
- The statement 'next gen" is "now gen" is somewhat true, but there is still a lot to learn about specific applications.
- There is a great need to continue to develop and pursue projects that will drive the implementation of the new technologies.
- These disruptive new technologies will affect the way in which biomedical research questions are addressed.

In response to Council member questions, Dr. Wilson made the following additional points:

- The number of staff at the Washington University Genome Sequencing Center is being decreased to accommodate reduction in the number of ABI 3730 instruments as production shifts to the new technologies.
- Paired ends have substantial value, work well in the Roche/454 driven system, and have very recently started to work in the Illumina/Solexa system.
- Portions of the genome that were previously undetectable because they are unclonable are being picked up by 454 sequencing.
- Transcriptome analysis will be a key to future research.

Whole exome sequencing

Dr. Richard Gibbs presented an overview of the analysis of regional variation through whole exome studies and next generation sequencing technology implementation at the Baylor Human Genome Sequencing Center. Similar to Washington University, Baylor has been decreasing the number of ABI 3730 in operation. Baylor is highly enthusiastic about the Roche/454 FLX technology because of their experience with microbial genomes. They have started implementing the technology on the baboon genome sequencing project using pooled BAC clones. The ABI SOLiD system has also already provided some good data, although it has long cycle times. Error rates are low with all

the new instruments and there is good complementarity of the types of errors from each platform. ABI 3730 sequencing would cost on the order of $400K per person for full exonic sequencing. The Nimblegen capture array technology can increase efficiency in this process; the implementation of this technology is moving along well, although there are still problems in obtaining even coverage across the genome. Capture technology will be important in the near term, until whole genome sequencing becomes routine.

Dr. Gibbs commented that Baylor has been engaged in many small resequencing projects to address human health. These small individual projects require a great deal of overhead. He argued that sequencing centers should focus on larger, comprehensive projects that enable technology.

De novo whole genome sequencing

Dr. Eric Lander presented an overview of *de novo* whole genome sequencing and next generation sequencing technology implementation at the Broad Institute. He presented rough cost estimates for sequencing technologies at Broad; the cost per gigabase of sequence for ABI 3730, 454 FLX and Solexa is close to $450,000, $25,000-$100,000 and $2,500-$10,000 respectively. Dr. Lander stated that continuing investment is needed in developing and implementing the new technologies because the machines are not operating at their limits yet. There is an enormous amount of work that needs to be invested in developing coverage models, new hybrid sequencing strategies, assembly methods and understanding errors rates. Although progress has already been made in costs and accuracy of the new technologies, the challenge is to make the data from the new technologies as high quality as the data from the retiring ABI 3730 machines.

Summary

Dr. Lander then proceeded to summarize the morning's scientific discussion. He emphasized that the new sequencing technologies have created enormous scientific opportunities, but we are still in a time of transition. The role of the large-scale sequencing centers is to develop the technologies to the point that they can be used effectively and efficiently. In the future, sequencing approaches will be used in four basic ways: signature sequencing, whole-genome resequencing, directed sequencing and whole-genome de novo assembly. There are proof-of-principle studies within each of these categories. Proof of principle, however, is not the same as production level. Over the next year, issues involving the sequencing instruments, sample preparation, data analysis and process will have to be addressed before the new technologies can be fully scaled up to production levels. The goal for the next year should be to develop a production-ready process and initial cost models.

Staff Proposal

Dr. Jane Peterson then presented the NHGRI staff proposal to adjust the emphasis of the large-scale sequencing program over the next 12 months to emphasize the implementation of the new sequencing technologies. The projects discussed this morning

are well suited to promote such implementation. One consequence of this reprioritization is that production for 2008 will be very diverse, with a more limited investment in individual organism sequencing. Only 50% of the program will be comprised of a collection of projects as they have been defined in the past. The program will de-emphasize producing finished, highly refined small projects in the short term, with the expectation that the payoff will be in terms of sequence production capacity and cost in the longer term. Another consequence that needs to be anticipated is that much of the data from the development activities will be of unknown quality.

If approved by Council, program staff will discuss the details of their individual proposals with each center. The program will be allowed as much flexibility as possible in expenditure of funds for implementation. The centers have agreed that they will be ready to discuss metrics for reporting production and cost data produced with new technologies at the July 2008 Steering Committee meeting and have regular reporting by the end of the year.

Council members' discussion focused on the projected change in production capacity. Dr. Adam Felsenfeld estimated that there could be as much as a 10 to 20-fold increase in capacity, even though the current production level pipeline will be cut in half. Dr. Lander reminded Council that we do not yet know what the quality of this increased capacity will be.

Sequencing Advisory Panel Comments

Dr. Rick Lifton summarized the Sequencing Advisory Panel's (SAP) thoughts about the staff proposal. He reviewed the progress that had been made in sequencing since the beginning of the Human Genome Project, and characterized it as "truly staggering." The entire field of comparative genomics has come out of the ability to generate huge amounts of sequence data, and this has fundamentally changed the way we think about biology, particularly human biology.

What has made the project work was the dedicated staff and the sequencing centers' understanding that they were holders of the public trust and needed to be transparent in what they were doing. Until now, the program has had a fairly linear path forward, but that will no longer be so. The new sequencing technologies have increased the complexity of applications and projects and the complexity of the management of the project will become much more difficult. From the Sequencing Advisors' point of view, the challenge will be how a small number of staff will be able to manage what will be an increasingly complex project. Maintaining accountability through this process will be essential, but difficult due to the transition. To maintain the public trust and interest, NHGRI will need to focus on those applications that are important to the constituencies that support us, i.e. those that address health interests and other public concerns. The program will need to find a balance between development and output.

Council Discussion

The Council then discussed the overall presentation and program proposal. The following points and questions were discussed:

- Public accessibility to the data has been very important to the program in the past and should continue as a priority.
- Should the centers become more development facilities and less production facilities in the future? Small projects can be now done in smaller laboratories; the challenge is to disseminate the technology. However, new challenges will arise that will require large-scale sequencing. Furthermore, projects like the 1000 Genomes are actually still relatively small compared to what will be needed to address many medical issues, i.e. analysis of samples from thousands of subjects in multiple different phenotypes. These types of projects will necessitate large-scale sequencing centers in the future.
- In this early phase of transition to the new technologies, one of the criteria being used to select projects are error issues and the robustness of the projects to error rates.
- The centers argued that 50 percent of the sequencing program is an appropriate level for new technology implementation and the concomitant need to cross-train staff on the technologies. The centers are prepared to make the staffing adjustments that will be a consequence of the introduction of new production pipelines.
- Functional elements and conserved regions should be included in the early studies, not just exons.
- Finishing remains a serious issue; to date, only two mammalian genomes have been finished to the high quality desired. The centers have discussed whether to address finishing issues during the twelve-month period under discussion, but decided they couldn't or shouldn't address it yet until we are much further along in implementing the new technologies in production.
- Paired end protocols for 454 and Solexa are coming along, but are not quite there yet. Other major bottlenecks to production are informatics, data analysis, and sample acquisition, particularly adequate consent for whole genome studies and data release.
- Dr. Collins noted that the new NIH GWAS policies are consistent with the approaches to data release that NHGRI needs for sequencing, so some progress is being made in this regard. But more needs to be done.
- Education of both the general public and members of the research community is important for success in the next phase of the large-scale sequencing centers and medical studies.

Dr. David Page summarized the Council discussion by noting that what has been proposed so far has addressed the "easy part" and that the issues that will confront us a year from now will be even harder to address. He suggested that finishing, informatics, appropriate consent and public education are issues that Council should discuss in upcoming sessions.

The Council voted to approve the Staff proposal to implement a reprioritization of the NHGRI large-scale sequencing program for the next twelve months.

DIRECTOR'S REPORT

I. GENERAL ANNOUNCEMENTS

NACGHR member Lincoln Stein is leaving Council as he has taken a new position as Platform Leader for Informatics and Biocomputing at the Ontario Institute for Cancer Research in Toronto. Dr. Stein will continue to be involved in the bioinformatics for the international cancer genomics effort and in modENCODE.

Dr. Zerhouni announced the appointment of Josephine Briggs, M.D., as Director of the National Center for Complementary and Alternative Medicine (NCCAM), effective February 3, 2008. Dr. Briggs has most recently served as a senior scientific officer at the Howard Hughes Medical Institute. She directed NIDDK's Division of Kidney, Urologic, and Hematologic Diseases from 1997 to 2006. Her scientific expertise, managerial and leadership talents focus on translational research, and understanding of the NIH all will serve her well in leading NCCAM.

On October 7th, the Nobel Foundation announced that the Nobel Prize in Physiology or Medicine for 2007 was being awarded jointly to Mario R. Capecchi, Martin J. Evans and Oliver Smithies for their discoveries of "principles for introducing specific gene modifications in mice by the use of embryonic stem cells."

On January 18th 2008, the Science and Technology Foundation of Japan announced that Victor A. McKusick, M.D., University Professor of Medical Genetics at the Johns Hopkins University School of Medicine, would be the 2008 recipient of the prestigious Japan Prize in Medical Genetics and Genomics, The sole laureate in his category this year, Dr. McKusick, who is widely renowned as the "father of genetic medicine," will receive a medal and 50 million yen ($470,000).

2008 electees to the Institute of Medicine include NHGRI-funded investigator Mike Boehnke of the University of Michigan; Andy Feinberg, the PI of an NHGRI CEGS at Johns Hopkins University; Aravinda Chakravarti also of Johns Hopkins, who is the current president of the ASHG and a former member of the NACHGR; and Wylie Burke of the University of Washington, who is an ELSI CEER PI and also a former member of the NACHGR.

Pamela L. Schwartzberg, M.D., Ph.D., senior investigator and head of the Cell Signaling Section of the Genetic Disease Research Branch of the National Human Genome Research Institute (NHGRI), will receive the AAI-BD Biosciences Investigator Award from the American Association of Immunologists (AAI). This award, announced January 25, 2008, recognizes Dr. Schwartzberg's outstanding, early-career research contributions to the field of immunology. The award recognizes Dr. Schwartzberg's high-impact observations in the broad field of T-cell signaling.

Paul P. Liu, M.D., Ph.D., senior investigator, Genetics and Molecular Biology Branch, NHGRI, received the 2008 Outstanding Achievement Award from the NIH Asian and

Pacific Islander American Organization (APAO). The award recognizes Dr. Liu's outstanding contributions to cancer research, signal transduction and developmental biology at NHGRI.

The journal *Science* named human genetic variation, an area of research pioneered by NHGRI-led efforts, as the top scientific breakthrough of 2007. As DNA sequencing technology becomes faster and cheaper, researchers are finding out exactly what makes individuals different from one another in ways not possible before. Techniques that scan for genetic differences are linking particular variations to traits and diseases at an astounding rate. In addition to human genetic variation, *Science* named the Human Microbiome, another NIH-led effort, as an "area to watch," along with microRNAs, synthetic genomes, and neural circuits.

It is time for NHGRI to begin a new planning process, as it has been five years since the publication of the "Vision" document in *Nature*. A new planning effort will ensure that NHGRI plans optimally for its own future within the context of the future of the wider field of genomics. This process is likely to be different from the last one; because much of the previous plan is still relevant and because of current budgetary and staffing constraints, there are not likely to be as many workshops and conferences this time. The new process will culminate with the submission to Council of a draft plan at the February 2010 meeting, with subsequent publication in a leading scientific journal. Council approved the new planning process.

II. NEW NHGRI INITIATIVES

Several ELSI program announcements have been reissued. They do not have funds set aside, but inform investigators about the areas in which the Institute is most interested currently.

NHLBI issued a Request for Applications titled "Development and Application of New Technologies to Targeted Genome-wide Resequencing in Well-Phenotyped Populations." NHGRI was a co-sponsor of that RFA. The receipt date was February 28, 2008.

RFAs have also recently been issued to solicit applications to several Roadmap programs, including the Human Microbiome Project, Epigenomics Project, Molecular Libraries Initiative and the NIH Director's Pioneer Award.

III. RECENT SCIENTIFIC ACCOMPLISHMENTS AND ISSUES

NHGRI - EXTRAMURAL PROGRAM

The *Drosophila* Comparative Genome Sequencing and Analysis Consortium published, on November 8, 2007, a comparative analysis of twelve fruit fly species, 10 of which were sequenced recently. The sequencing was done at Agencourt Biosciences, Washington University, the Broad Institute, and the J. Craig Venter Institute with NHGRI funding. The availability of such a high number of closely related species allows for higher resolution and increased power of comparative analyses.

The NHGRI Research Network for Large-scale Sequencing and the NHGRI Sequencing Advisory Panel met on December 4-5, 2007 at the Washington University Genome Sequencing Center in St. Louis, Missouri.

On January 22nd, NHGRI announced the 1000 Genomes Project, an ambitious effort that will involve sequencing the genomes of at least a thousand people from around the world to create the most detailed and medically useful picture to date of human genetic variation. The project will also receive major support from the Wellcome Trust Sanger Institute in Hinxton, England, and the Beijing Genomics Institute, Shenzhen (BGI Shenzhen) in China. The scientific goals of the 1000 Genomes Project are to produce a catalog of variants that are present at 1 percent or greater frequency in the human population across most of the genome, and down to 0.5 percent or lower within genes.

Relevant to the morning's discussion of new sequencing technologies, Council was given a brief status report on commercialization of new sequencing technologies, many of which had received partial grant support from NHGRI during development.

The International HapMap Consortium published analyses of its second-generation map of human genetic variation, which contains more than 3.1 million SNPs. In two papers in the journal *Nature*, the Consortium described how the higher resolution map offers greater power to detect genetic variants involved in common diseases, to explore the structure of human genetic variation, and to learn how environmental factors have shaped the human genome. Genome-wide association studies have already given researchers new insights into the biological basis of many diseases. The rapid growth of genome-wide association studies over the past year-and-a-half has been fueled by the HapMap consortium's decision to make its SNP datasets immediately available in public databases. The continued refinement of the haplotype map, which is planned as HapMap Phase III, means genetic association discoveries will come even faster, easier, and cheaper.

In November 2007, TSP researchers announced the discovery of 57 genomic changes that occur frequently in lung cancer patients. More than 40 of these appear to be associated with genes not previously known to be involved in lung adenocarcinoma. TSP has provided a genomic landscape of lung cancer that has already implicated a novel gene in controlling lung cell growth. Moreover, the findings may suggest new ways of attacking this deadly cancer.

In September 2007, the NHGRI announced grants totaling more than $80 million over the next four years to expand the ENCyclopedia Of DNA Elements (ENCODE) project, which in its pilot phase yielded provocative new insights into the organization and function of the human genome. In June, the ENCODE research consortium published a set of landmark papers in the journals *Nature* and *Genome Research* that found the organization, function and evolution of the genome to be far more complicated than most had suspected. An ENCODE Consortium meeting on November 28-29, 2007 brought together members of the ENCODE Consortium, the External Consultants Panel,

modENCODE PIs, ENCODE Technology Development PIs and other investigators who are working on ENCODE-related projects. Consortium members discussed lessons learned from the ENCODE pilot project, common resources to be used by investigators, data release and publication, progress reporting, and the application of sequencing technologies to methods being used in ENCODE. A joint meeting between the two Consortia is planned for June.

Applications for the ENCODE Data Analysis Center (RFA HG-07-010, which solicited applications for a resource to work with the ENCODE Analysis Working Group and the ENCODE Data Coordination Center to coordinate, support, and assist in the analysis of data produced by the ENCODE Consortium) will be considered during the Closed Session of this Council meeting.

The status of the Mammalian Gene Collection (MGC), Xenopus Gene Collection (XGC), and Zebrafish Gene Collection (ZGC) were reported. The last human and mouse cDNA clones for the Mammalian Gene Collection (MGC) are scheduled for delivery by June 2008, with projections of verified clones for at least 94% and 91% of the well-supported (NM accession) human and mouse genes. DNA synthesis is being used to generate clones for as many as possible of the genes that are still not represented in MGC. To date, synthetic clones for 2,000 human and mouse genes have been delivered, and another 1,300 are in progress. The C list MGC program to predict and verify new human genes has identified 563 novel gene sequences. Over 160 are from previously unreported genes, with most of the remainder significantly expanding sequences of known genes. The MGC, ZGC, and XGC clones, previously archived within the IMAGE Consortium at Lawrence Livermore National Laboratory, have been recently relocated to a permanent new home at the HudsonAlpha Institute for Biotechnology, in Huntsville, AL.

In addition to human cells, the body is colonized by large numbers of microorganisms. Some of these are critical to the maintenance of human health, others cause illness. To better understand the biology of the human microflora and its relationship to human health and disease, the NIH has launched the Human Microbiome Project (HMP) as part of the Roadmap for Medical Research and has committed $115 million over the next five years to it. The HMP will utilize the latest sequencing technology and other tools to determine the feasibility and value of exploring the complex relationships between the microbiome and human health. The NIH office that administers the Roadmap (the Office of Portfolio Analysis and Strategic Initiatives, OPASI) provided FY07 funds to "jumpstart" the HMP. These funds were awarded to the Genome Sequencing Centers at the Baylor College of Medicine, the Broad Institute, the J. Craig Venter Institute and the Washington University School of Medicine to begin work on Initiative 1 of the HMP, the construction of a reference data set of complete genome sequences of 600 microbes isolated from five body sites (the gut, oral cavity, skin, vagina, and nasal passages).

HMP is planning a workshop to be held in the Washington DC area on March 27, 2008 that will be co-chaired by Jane Peterson (NIH/NHGRI) and George Weinstock (Washington University School of Medicine). The workshop is designed to seek input on ways to engage members of the larger scientific community by leveraging the knowledge

of the research community to ensure that the jumpstart phase of the HMP, as well as the entire HMP program, builds upon existing knowledge in the field.

Another new Roadmap effort is the Epigenomics program. Epigenomics is an emerging frontier of science that involves the study of inheritable phenotypes that are not determined by DNA sequence. The Roadmap Epigenomics Program proposes to: (1) develop a set of maps of reference epigenomes; (2) develop standardized platforms, procedures, and reagents for epigenomics research; (3) conduct demonstration projects to study the epigenomic determination of health effects; (4) develop new technologies for single cell epigenomic analysis and in vivo imaging of epigenetic activity; and (5) create a public data resource to accelerate the application of epigenomic approaches. This program is co-led by NIDA and NIEHS.

A workshop is planned to provide guidance to the NIH on potential computational approaches to maximizing the knowledge and value derived from the vast amount of data being generated by the Molecular Libraries Initiative (MLI) and to identify specific community needs for improved informatics and cheminformatic tools to fully develop and explore these approaches. This new field of systems chemical biology is in early development and there is a need for improved investigative tools. The focus of this workshop will be on exploring the potential applications for the MLI data, on determining the best ways to make these complex data available to the scientific community, and to identify novel software tools that will be required to extract the full potential from the data.

In October, NHGRI announced the establishment of two new Centers of Excellence in ELSI Research (CEERS) to address critical ethical, legal and social questions faced by genetic and genomic research. The new center grants were awarded to at the University of North Carolina-Chapel Hill (Gail Henderson, P.I.) and the University of Pennsylvania (Reed Pyeritz, P.I.). .

The ELSI Assessment Panel, a working group of Council chaired by Harold Shapiro, has now met twice face-to-face and had a number of meetings via teleconference. The Panel has received written input from numerous individuals from a variety of backgrounds and perspectives and, on February 1, interviewed about two dozen individuals, also with a variety of perspectives, in person to get their input. The Panel expects to present its report to Council at the May meeting.

NHGRI – INTRAMURAL PROGRAM

The NIH has created a new intramural center, focusing on genomics and health disparities. The mission of the NIH Intramural Center for Genomics and Health Disparities (NICGHD) is to advance research into the role of culture, lifestyle, genetics and genomics in health disparities. Dr. Charles Rotimi has been appointed as the Director of the Center. He will guide the development of genetic epidemiology models and population genetics research for exploring the patterns and determinants of common complex diseases in populations of the African Diaspora and other human populations.

Both environmental and genetic factors influence a person's blood fat, or lipid levels, an important risk factor for coronary artery disease (CAD). An international collaboration supported primarily by the NIH, has discovered more than 25 genetic variants in 18 genes connected to cholesterol and lipid levels. These variants potentially open the door to strategies for the treatment and prevention of CAD. In an equally interesting finding, an international team, supported in part by NHGRI, has found evidence that common genetic variants recently linked to osteoarthritis may also play a minor role in the determination of human height. Both studies illustrate the continuing emergence of genetic findings through teamwork and NIH leadership, as well as direct clinical relevance of genetic research.

Premature birth is a major cause of infant death and increases the risk of many potentially disabling conditions. A surprising new study, published in October 2007 and led by NHGRI intramural researchers, showed that pregnant women who have very low cholesterol levels might face a greater risk of delivering their babies prematurely. Researchers noted a differential impact of low cholesterol levels on the rates of premature delivery in white and African-American mothers, which is of particular interest since premature delivery is a leading cause of health disparities. These findings may lead to a better understanding of the biology of birth defects and, eventually, suggest more effective strategies for preventing them.

According to NIH procedures, an external group is reviewing the performance of Dr. Eric Green as the Scientific Director of the NHGRI Intramural Research Program. The committee's report will be presented to Council at it May meeting.

NHGRI OFFICE OF THE DIRECTOR

The NIH has selected the first projects to be funded as part of the Genes, Environment and Health Initiative (GEI), a unique collaboration between geneticists and environmental scientists. These projects are part of a broader effort across HHS agencies to build on recent advances in genomic science and medicine, including the Secretary's Initiative on Personalized Health Care.

The cooperative agreement for Genome-Wide Studies in Biorepositories with Electronic Medical Record Data, dubbed the "eMERGE Network" (Electronic Medical Records and Genomics Network), was awarded in September, with co-funding from the NHGRI ELSI program and additional support from NIGMS.

A workshop on "Frontiers in Population Genomics: Research Directions for NHGRI" was held December 18-19 to review research progress in population genomics and to propose ideas for research directions for the next 3-5 years. Presentations and other materials from the workshop will be available soon on the Office of Population Genomics website at http://www.genome.gov/19518660 and a report will be presented at the May Council meeting.

As discussed at the September Council meeting, the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies was released in August 2007, and went into effect with applications submitted for the receipt date of January 25, 2008. On November 16th, 2007, a supplemental notice with implementation guidance on the policy for NIH applicants (and their institutions) was released by the Office of Extramural Research. Included within this Notice were instructions regarding where and how within an application to describe the plans for submitting GWAS data to the NIH central repository, dbGaP, as well as how to indicate plans for obtaining access to GWAS datasets within dbGaP if relevant to a proposed project. In addition to the new instructions for applicants, the NIH has posted a series of FAQs on the policy and expected procedures for investigators and institutions, and a Points to Consider document for IRBs and Institutions to help inform the review of datasets as required by the policy prior to submission to dbGaP. Also, the past few months have seen the launch of the trans-NIH governance system that will provide on-going oversight and management of the policy's implementation across the NIH. New developments and additional information for investigators will be posted to the GWAS homepage as materials are developed.

As noted in the policy, data access and use for each NIH GWAS dataset will be overseen by a Data Access Committee (DAC) to evaluate investigators' requests for access to the datasets. NHGRI will use the same model for access to data from large-scale medical sequencing studies. The GAIN DAC has been reviewing requests for access to the GAIN genome-wide association datasets since May 2007, and has pioneered many of the procedures and documents now being utilized across NIH. The NHGRI DAC was formed to review requests for data from the medical sequencing projects and recently has been charged with responsibilities for GWAS datasets that are in dbGaP but are not overseen by the DAC of any other IC, as well those datasets that are part of NHGRI-led initiatives. The Cancer Genome Atlas project has human subjects privacy concerns akin to those of GWAS and has adopted policies and procedures consistent with the NIH GWAS Policy. NHGRI and NCI have established a DAC for evaluation of requests to access TCGA individual-level datasets. The actions of this and the other DACs will be monitored by the Working Group to the Advisory Committee to the NIH Director.

In response to questions and comments from Council members, the following points were made:

- There are a couple datasets in dbGaP in which the participants provided consent only for non-commercial use, but the vast majority of data is available for all users.
- Tracking publication behavior and tracking data security has been an issue for all of the NIH studies that have provided GWAS data. There is one working committee of the DAC chairs attempting to develop a uniform policy for this purpose, and another that is trying to develop ways to monitor use of the data. Dr. Zerhouni has appointed a group to try to achieve a balance between access to data and privacy.
- Acknowledging that it will be hard to track IP (because you won't know about it until the IP is published), the best that NIH can do is to deposit the data publicly to establish prior art. One Council member noted that eventually a way to police or monitor this might be needed.

The Trans-NIH Communications Group on Genetics and Common Disease is charged with rapidly developing and implementing a cohesive communications plan to inform and educate both the public and health professionals about the genetics of common disease and traits. The recent outpouring of data from genome-wide association studies should rapidly advance scientific understanding of the genetics of common disease and also enable new avenues of basic and applied research. These data have, however, also created opportunities for new companies, such as 23andMe, deCODEme, and Navigenics to market personalized genetic information to consumers directly, via the web. Both the public and health professionals are currently ill-prepared to deal with the health implications of this increasingly abundant information or to deal with its commercial use, and there is a danger of overstating the immediate benefits for clinical medicine. NIH is trying to move quickly and plans to create a rapid response team that has expertise in the area but also develops specific talking points. There will be a more in-depth discussion of this at a later Council meeting. In addition, the ELSI discussion panel at the 2008 Biology of Genomes meeting, to be held at Cold Spring Harbor in May, will address the topic.

In 2004, the NHGRI hosted a successful roundtable on Race, Ethnicity, and Genetics. To further the work accomplished at this roundtable, the NHGRI plans to host an invitational retreat summit in 2008 on Race, Ethnicity, Genetics and Health Disparities. The objective of the meeting will be to bring together researchers from diverse fields to discuss what we know and don't know about genomics and racial and ethnic health disparities. The summit will be co-sponsored by the National Center for Minority Heath and Health Disparities and the National Cancer Institute.

A carrier screening conference was held on February 6-7, 2008. It engaged a variety of organizations in a discussion of emerging opportunities and obstacles in carrier screening for single gene disorders. Co-sponsors included the National Institute of Child Health and Human Development (NICHD), the Office of Rare Diseases (ORD), the Health Resources and Services Administration (HRSA), the Centers for Disease Control (CDC), the Genetic Alliance and the American College of Medical Genetics (ACMG).

A new online community for genetics teachers to exchange learning resources and techniques, and to participate in NHGRI educational events was officially launched on December 3, 2007. In the first month, membership reached 125 and is predicted to reach 300 members in the first year. The community was formally announced in a keynote lecture by NHGRI Director Francis Collins at the national meeting of the National Association of Biology Teachers (NABT). The "Community of Genetic Educators" is based on social computing techniques and is a collaborative website, one of the first of this nature in federal government

NHGRI – POLICY

After a protracted dispute with the White House about overall funding levels for domestic programs, and a sustained veto of the first version of the Labor-HHS spending bill,

Congress passed the Labor-HHS appropriations bill in late December. The numbers are disappointing for the NIH, with the total appropriation of $29,228,541,000 representing just a 1.1% increase from the FY 2007 enacted level. NHGRI's FY 2008 budget is $486,779,000 which represents only a 0.1% increase from the FY 2007 enacted funding level. With a net increase of only $288,000, the Institute will be severely constrained in what we can endeavor to fund in 2008. The President's Budget request for FY 2009 was released last week and it does not change the outlook much for NHGRI. The NIH budget is now close to twenty percent down in buying power from the budget at the end of the doubling in 2003 because of inflation.

The Genetic Information Non-Discrimination Act (GINA) is still stuck in the Senate. Senator Coburn has maintained his hold on the bill, which is still preventing the bill from being moved to the floor. There was an effort mounted in December by GINA advocates to get the bill attached to the omnibus spending bill but, despite cooperation from House and Senate appropriators, the attachment was not permitted by the Senate leadership. Congressional staff are continuing to work on moving the bill and discussing strategies with the Senate leadership.

Last summer, Senator Edward Kennedy introduced the Minority Health Improvement and Health Disparity Elimination Act, cosponsored by Senators Cochran, Obama, Clinton, Bingaman, Brown and Durbin As currently written, the Act would require the Director of NHGRI to convene a summit on race, genetics, health and environment, and to issue a report with specific recommendations and guidelines. In early January, Vence Bonham and M.K. Holohan of NHGRI met with Senator Kennedy's HELP committee staff to discuss NHGRI's ongoing work in relation to genetic variation, race, and health disparities, as well as NHGRI's ongoing planning for a summit on race and genetics in the spring of 2008. NHGRI will continue discussions with the HELP Committee staff, particularly in regard to the agenda and goals of the NHGRI summit, and to explore ways in which that process could inform the Committee's plans for the completion of the Act.

The Secretary's Advisory Committee on Genetics, Health and Society (SACGHS) will meet on February 12-13, 2008 at the Humphrey building in Washington, DC. The primary focus of this meeting will be to finalize the oversight recommendations in the Draft Report on the Oversight of Genetic Testing, which was commissioned by Secretary Leavitt. The recommendations will be sent to the Secretary by the end of February (along with a draft of the report) and the final draft of the oversight report will be sent to the Secretary by the end of April.

Also at the meeting, Dr. Steven Teutch will take over from Dr. Reed Tuckson as the SACGHS chair. Dr Teutch is the Executive Director for Outcomes Research & Management at Merck & Co., Inc., and has been a member of the committee since 2006.

Dr. Guyer announced that Francis Collins was awarded the Presidential Medal of Freedom.

COUNCIL-INITIATED DISCUSSION

The ELSI Advisory Panel report, a report of the review of the Scientific Director and a report from the Frontiers in Population Genomics meeting will be on the agenda for the May 2008 Council meeting.

Dr. Richard Gibbs suggested discussing informatics as one of the four bottlenecks to new technology implementation.

STATEMENT OF UNDERSTANDING BETWEEN COUNCIL AND NHGRI

Dr. Guyer noted that this statement is read and agreed upon each year. The only change from last year is an item that mentions a cap on supplemental awards made without Council approval. The Statement of Understanding was adopted unanimously.

CONFLICT OF INTEREST

Dr. Guyer read the Conflict of Interest policy to Council and reminded the members to sign the forms provided.

REVIEW OF APPLICATIONS

I hereby certify that, to the best of my knowledge, the foregoing minutes are accurate and complete.

_____ _____

Date

Mark Guyer, Ph.D.
Executive Secretary
National Advisory Council for Human Genome Research

_____ _____

Date

Francis S. Collins, M.D., Ph.D.
Chairman
National Advisory Council for Human Genome Research