Future Opportunities for Genome Sequencing… and Beyond

Welcome, Charge, and Context

Eric Green, M.D., Ph.D.
Director, NHGRI

NIH National Human Genome Research Institute


Thank You
Asante
Shukria
köszönöm
Vinaka
Kiitos
Maake
Dhanyavadagalu
Dankon
Manana
Kasih
Matondo
감사합니다
Kam Sah Hammida
спасибо
Dankscheen
Dank Je
Mauruuru
Biyan
Terima
Taiku
Tack
Blagodaram
Ngiyabonga
Dziekuje
Chokrane
Diolch i Chi
Grazie
Juspaxar
Arigato
Mochchakkeram
Gracias
Paldies
Tingki
நன்றி
धन्यवाद
Ďakujem
Bedankt
Grazas
cảm ơn bạn
Khap
Gratias Tibi
Ua Tsaug Rau Koj
Di Ou Mèsi
Kia Ora
Obrigado
Děkuji
Nirringrazzjak
Kop Khun
ありがとう
Suksama
Hvala
Djiere Dieuf
Rahmat
Nuwun
Matur
谢
Welalin
Misaotra
Danke
Go Raibh Maith Agat
Eskerrik Asko
хвала
Merci
ขอขอบคุณคุณ
Salamat
Najis Tuke

NHGRI: Guided by Strategic Planning

1991-1995

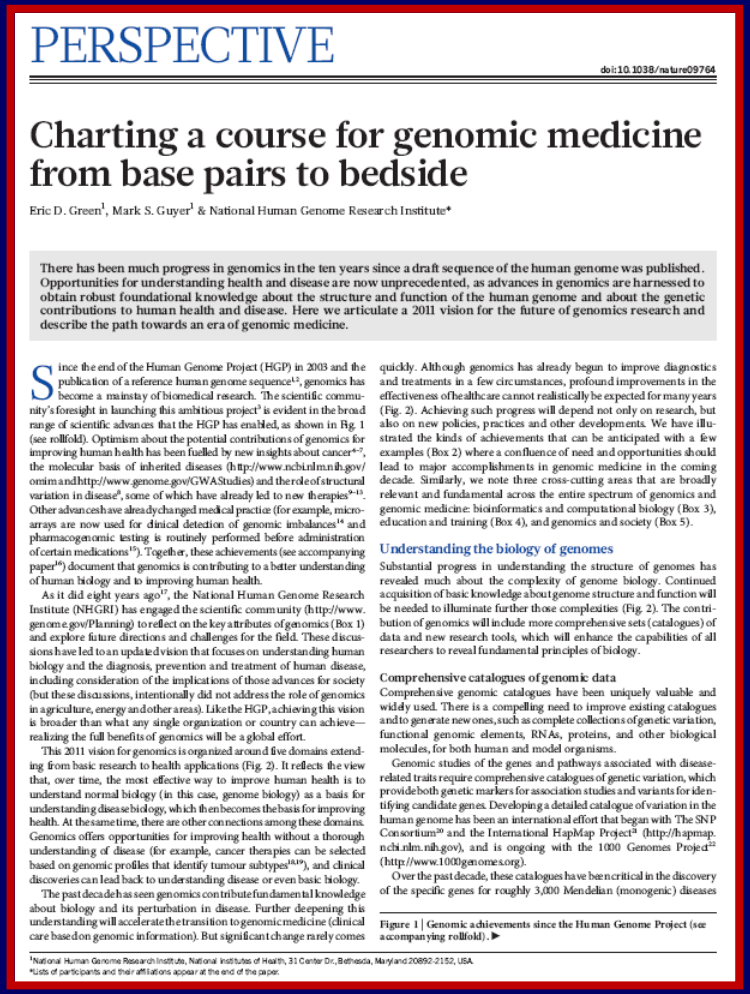
1993-1998

1998-2003

2003-2010

2011-Present

NHGRI's Current Strategic Plan



PERSPECTIVE

doi:10.1038/nature09764

Charting a course for genomic medicine from base pairs to bedside

Eric D. Green[1], Mark S. Guyer[1] & National Human Genome Research Institute*

There has been much progress in genomics in the ten years since a draft sequence of the human genome was published. Opportunities for understanding health and disease are now unprecedented, as advances in genomics are harnessed to obtain robust foundational knowledge about the structure and function of the human genome and about the genetic contributions to human health and disease. Here we articulate a 2011 vision for the future of genomics research and describe the path towards an era of genomic medicine.

Since the end of the Human Genome Project (HGP) in 2003 and the publication of a reference human genome sequence[1,2], genomics has become a mainstay of biomedical research. The scientific community's foresight in launching this ambitious project[3] is evident in the broad range of scientific advances that the HGP has enabled, as shown in Fig. 1 (see rollfold). Optimism about the potential contributions of genomics for improving human health has been fuelled by new insights about cancer[4–7], the molecular basis of inherited diseases (http://www.ncbi.nlm.nih.gov/omim and http://www.genome.gov/GWAStudies) and the role of structural variation in disease[8], some of which have already led to new therapies[9–13]. Other advances have already changed medical practice (for example, microarrays are now used for clinical detection of genomic imbalances[14] and pharmacogenomic testing is routinely performed before administration of certain medications[15]). Together, these achievements (see accompanying paper[16]) document that genomics is contributing to a better understanding of human biology and to improving human health.

As it did eight years ago[17], the National Human Genome Research Institute (NHGRI) has engaged the scientific community (http://www.genome.gov/Planning) to reflect on the key attributes of genomics (Box 1) and explore future directions and challenges for the field. These discussions have led to an updated vision that focuses on understanding human biology and the diagnosis, prevention and treatment of human disease, including consideration of the implications of those advances for society (but these discussions, intentionally did not address the role of genomics in agriculture, energy and other areas). Like the HGP, achieving this vision is broader than what any single organization or country can achieve—realizing the full benefits of genomics will be a global effort.

This 2011 vision for genomics is organized around five domains extending from basic research to health applications (Fig. 2). It reflects the view that, over time, the most effective way to improve human health is to understand normal biology (in this case, genome biology) as a basis for understanding disease biology, which then becomes the basis for improving health. At the same time, there are other connections among these domains. Genomics offers opportunities for improving health without a thorough understanding of disease (for example, cancer therapies can be selected based on genomic profiles that identify tumour subtypes[18,19]), and clinical discoveries can lead back to understanding disease or even basic biology.

The past decade has seen genomics contribute fundamental knowledge about biology and its perturbation in disease. Further deepening this understanding will accelerate the transition to genomic medicine (clinical care based on genomic information). But significant change rarely comes quickly. Although genomics has already begun to improve diagnostics and treatments in a few circumstances, profound improvements in the effectiveness of healthcare cannot realistically be expected for many years (Fig. 2). Achieving such progress will depend not only on research, but also on new policies, practices and other developments. We have illustrated the kinds of achievements that can be anticipated with a few examples (Box 2) where a confluence of need and opportunities should lead to major accomplishments in genomic medicine in the coming decade. Similarly, we note three cross-cutting areas that are broadly relevant and fundamental across the entire spectrum of genomics and genomic medicine: bioinformatics and computational biology (Box 3), education and training (Box 4), and genomics and society (Box 5).

Understanding the biology of genomes

Substantial progress in understanding the structure of genomes has revealed much about the complexity of genome biology. Continued acquisition of basic knowledge about genome structure and function will be needed to illuminate further those complexities (Fig. 2). The contribution of genomics will include more comprehensive sets (catalogues) of data and new research tools, which will enhance the capabilities of all researchers to reveal fundamental principles of biology.

Comprehensive catalogues of genomic data

Comprehensive genomic catalogues have been uniquely valuable and widely used. There is a compelling need to improve existing catalogues and to generate new ones, such as complete collections of genetic variation, functional genomic elements, RNAs, proteins, and other biological molecules, for both human and model organisms.

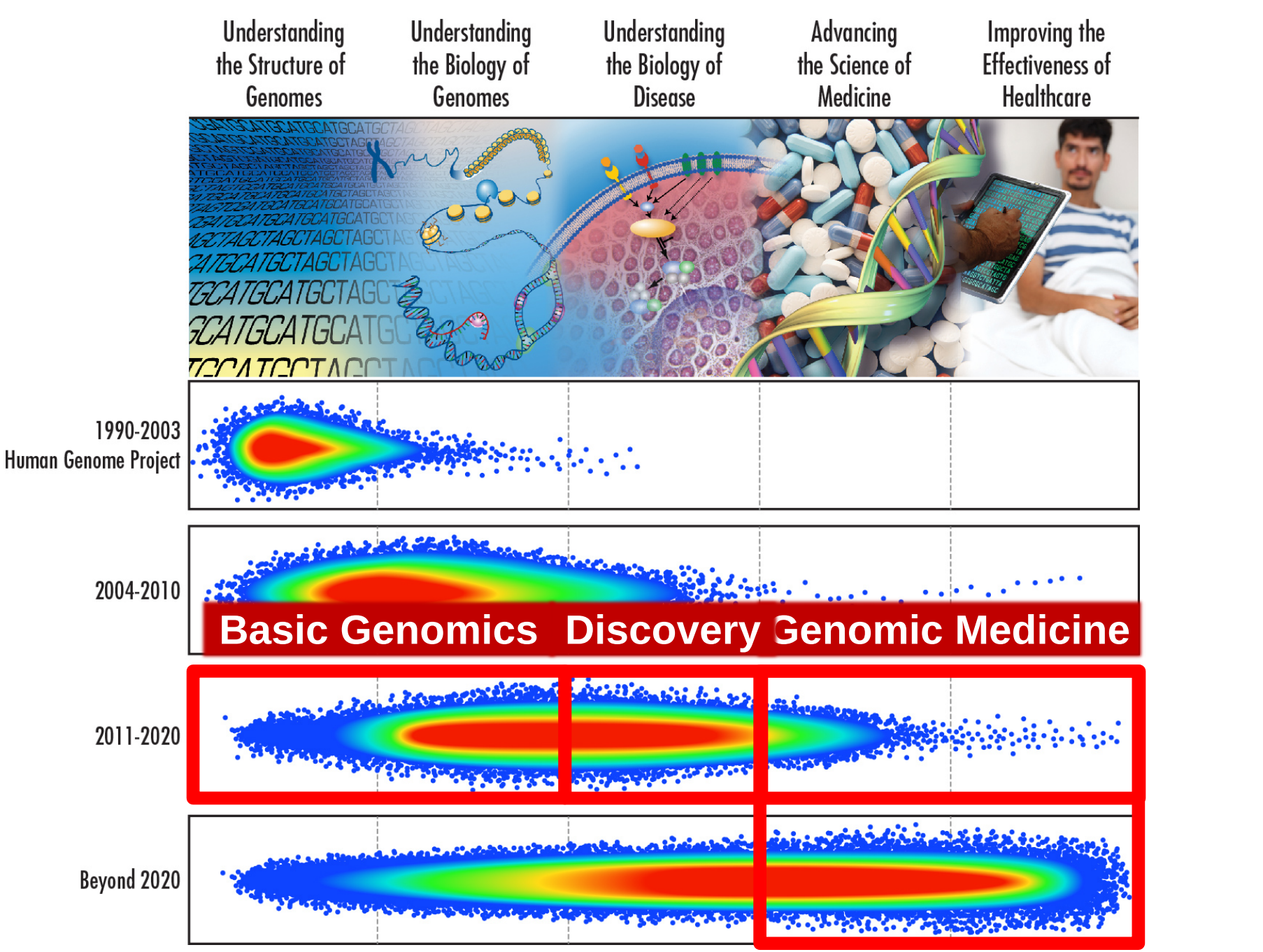
Genomic studies of the genes and pathways associated with disease-related traits require comprehensive catalogues of genetic variation, which provide both genetic markers for association studies and variants for identifying candidate genes. Developing a detailed catalogue of variation in the human genome has been an international effort that began with The SNP Consortium[20] and the International HapMap Project[21] (http://hapmap.ncbi.nlm.nih.gov), and is ongoing with the 1000 Genomes Project[22] (http://www.1000genomes.org).

Over the past decade, these catalogues have been critical in the discovery of the specific genes for roughly 3,000 Mendelian (monogenic) diseases

Figure 1 | Genomic achievements since the Human Genome Project (see accompanying rollfold). ▶

[1]National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., Bethesda, Maryland 20892-2152, USA.
*Lists of participants and their affiliations appear at the end of the paper.

February 2011



Green et al. 2011

NHGRI's 'Flagship':
The Genome Sequencing Program

1. History

2. Characteristics

NHGRI's 'Flagship': The Genome Sequencing Program

1. History

2. Characteristics

Post-HGP Planning

~2002

Program Review Workshop

2005

2003 2006 2012 2016

Human Genome Project

1990-2003

"Large-Scale Genome Sequencing Centers"

2003-2006

Comparative Genomics

HapMap Project

June 2005 Program Review Workshop

Workshop on the Future of the Large-Scale Sequencing Program

Share Print

June 13, 2005

Executive Summary

The National Human Genome Research Institute convened a workshop to obtain opinions from the scientific community on the current status and potential future directions of the NHGRI large-scale sequencing program. Participants were asked to consider the scientific, technological, and strategic opportunities in evaluating NHGRI's future investment in sequencing, and to specifically address several general questions and challenges:

- Given what has already been accomplished - very high quality assembled genome sequences of the human and major model organisms, draft sequence assemblies of genomes representing many of the nodes of the metazoan lineage, concerted application of comparative sequencing to annotate mammalian genomes - what are the best future opportunities for large-scale sequencing? What is the proper balance of these types of projects going forward? Should other kinds of large-scale sequencing projects be considered? What is the continuing priority of large-scale sequencing as a source of genomic data compared with other types of genomic data?

- Disruptive technologies appear to be promising enough that a significant reduction in the cost of DNA sequencing could occur within the next three years. What are the realistic prospects for the introduction of such a disruptive technology? How should it be anticipated and encouraged? How would it affect sequencing costs and capacity? How would it affect the types of scientific questions that can be addressed? How should the possibility of future significant cost reductions affect the decisions about the types of sequencing projects that should be initiated in the next two to three years?

- How should NHGRI evaluate the ongoing value of its investment in a large-scale sequencing program? How should it assess the contribution that continued sequencing will make to scientific research overall and genomic research in

genome.gov/16015129

Post-HGP Planning

~2002

Program Review Workshop

2005

Program Review Workshop

2009

2003 2006 2012 2016

Human Genome Project	"Large-Scale Genome Sequencing Centers"	"Large-Scale Genome Sequencing Centers"
1990-2003	2003-2006	2006-2011

Comparative Genomics

Microbiome

HapMap Project

1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing

Pathogens & Vectors

March 2009 Program Review Workshop

Workshop Report
The Future of DNA Sequencing at the National Human Genome Research Institute
March 23-24, 2009

What are the most important biomedical questions that can be addressed with large-scale sequence data? What are the most compelling sequence-based community resources that should be generated? What are the consequences of the rapid increase in sequencing capacity, and the rapid decrease in cost, afforded by the new technology platforms? In order to answer these questions, the National Human Genome Research Institute (NHGRI) convened a workshop to discuss the future of large-scale sequencing as one component[1] of the Institute's current two-year planning process for all of its scientific programs.

The need for this workshop was particularly underscored by the recent and ongoing rapid changes in sequencing technology, propelled by the "next generation" sequencing platforms. Introduced into production activities less than two years ago, the new sequencing platforms have already afforded an increase in throughput[2] of two orders of magnitude over the previous platforms, and this is likely to increase by nearly another order of magnitude in the next year or two. Furthermore, yet newer technologies are being developed and are expected to be available in the next three to five years. These rapid changes offer incredible new opportunities as well as major new challenges for the use of sequencing technology in general and to NHGRI's sequencing program specifically. As the technology continues to improve, new applications of genomic sequencing are constantly being developed, for example the sequencing of genomes from large numbers of individuals for disease and population studies, quantitative transcriptional analysis and epigenomics.

The 'disruptive' technological change has many other consequences. Most obviously, the ability to apply large-scale sequencing efficiently towards a larger number of problems will result in unprecedented demands on scientists' ability to find enough samples that are appropriate to addressing an expanded range of questions. To date, the most difficult problem has been obtaining samples for human disease or population studies that are properly consented for the work. One can also foresee

genome.gov/Pages/Research/DER/DERReportsPublications/SeqPlanWkshopReportFinal.pdf

Post-HGP Planning ~2002

Program Review Workshop 2005

Program Review Workshop 2009

2003 2006 2012 2016

Human Genome Project	"Large-Scale Genome Sequencing Centers"	"Large-Scale Genome Sequencing Centers"	4-Component "Genome Sequencing Program"
1990-2003	2003-2006	2006-2011	2012-2015

Comparative Genomics

HapMap Project

Microbiome

1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing

Pathogens & Vectors

NHGRI Genome Sequencing Program
Circa 2012-2015

Large-Scale Genome Sequencing and Analysis Centers

cser

Clinical Sequencing Exploratory Research

Post-HGP Planning

~2002

Program Review Workshop

2005

Program Review Workshop

2009

Program Review Workshop

2014

2003 2006 2012 2016

Human Genome Project	"Large-Scale Genome Sequencing Centers"	"Large-Scale Genome Sequencing Centers"	4-Component "Genome Sequencing Program"
1990-2003	2003-2006	2006-2011	2012-2015

Comparative Genomics

HapMap Project

Microbiome

1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing

Pathogens & Vectors

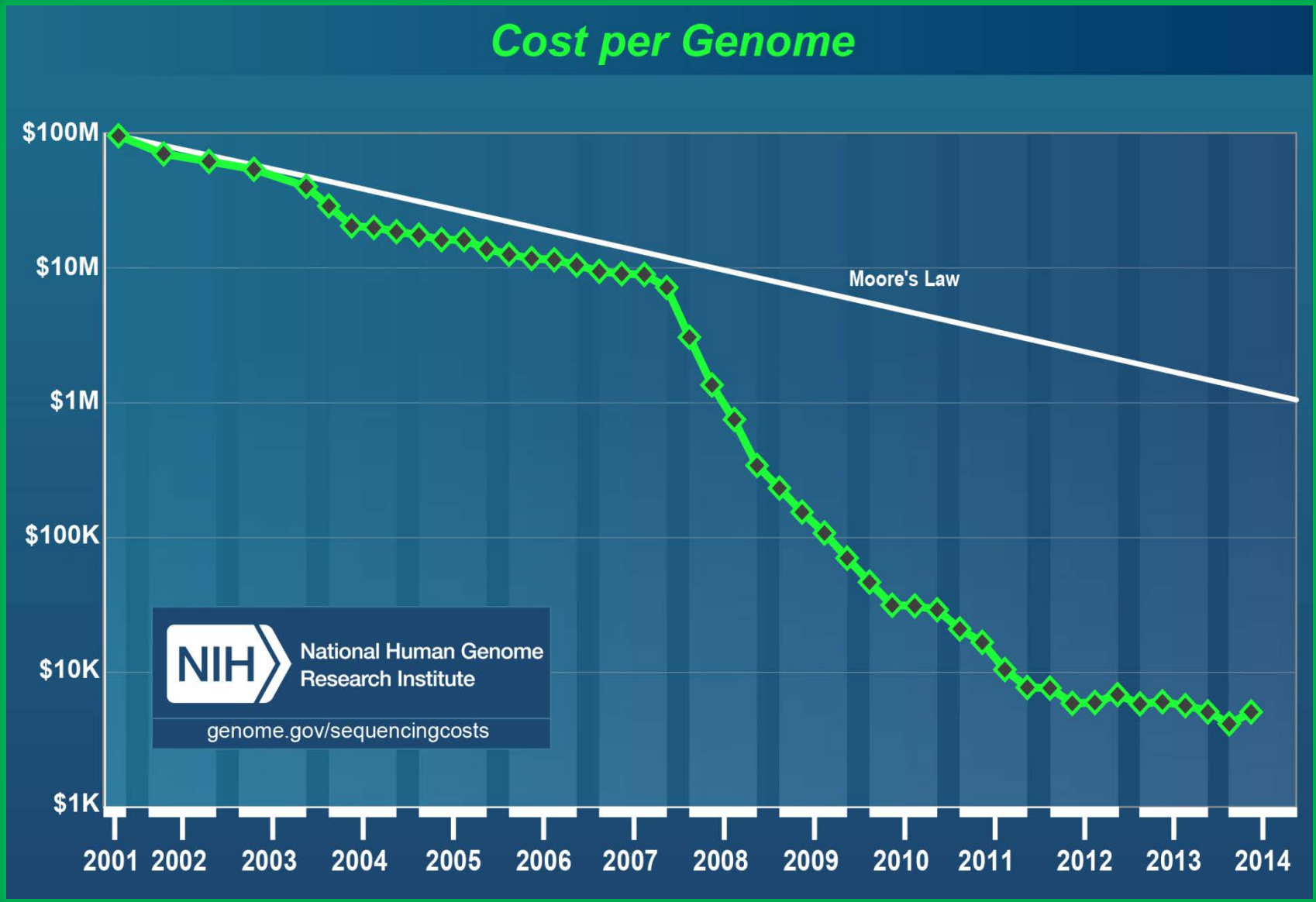
Common Diseases

Mendelian Diseases

Clinical Genomics

Computational Tools

Plummeting Cost of Genome Sequencing



Post-HGP Planning ~2002

Program Review Workshop 2005

Program Review Workshop 2009

Program Review Workshop 2014

2003 2006 2012 2016

Human Genome Project	"Large-Scale Genome Sequencing Centers"	"Large-Scale Genome Sequencing Centers"	4-Component "Genome Sequencing Program"	To Be Determined
1990-2003	2003-2006	2006-2011	2012-2015	2016-???

Comparative Genomics

HapMap Project

Microbiome

1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing

Pathogens & Vectors

Common Diseases ???

Mendelian Diseases ???

Clinical Genomics ???

Computational Tools ???

NHGRI's 'Flagship':
The Genome Sequencing Program

1. History

2. Characteristics

NHGRI Genome Sequencing Program: Characteristics (To Date)

- Large (i.e., Scale)
- Consortia-oriented
- Highly managed
- Resource-generating
- ‘Technology’-advancing
- Scientifically/medically relevant
- Nimble

Going Forward: What Does NHGRI Want?

- Continue being ‘genomics trailblazers’
- Alignment with strategic vision/plan
- Impact that correlates with program size
- If continuation of a major program, then retain 7 characteristics (previous slide)
- Importance of ‘moving on’ past initial catalytic role (e.g., organism sequencing, microbes, microbiome, and cancer)
- Increased ‘cost-sharing’ to broaden impact

NHGRI and Cost-Sharing

- Genomics = Huge; NHGRI = Small
- NHGRI cannot support ‘everything genomics’
- Partnerships are key (past and future)
- Consider formalizing an approach for cost-sharing in large-scale NHGRI-funded genomics projects

Purpose of Workshop

1. For Starters: It's what we do…
2. General: Natural time for strategic input (e.g., 3 years since 2011 strategic plan)
3. Critical: Synchronize strategic thinking in light of rapidly changing (and complicated) landscape
4. Practical: Fiscal Year 2016 and ~$100M (~25% of extramural funds)

Highest Priority: Discussion about Areas Associated with:

<u>Also</u> Important: Discussion about Areas Associated with:

Questions to Address (Among Many)

1. What are the ‘grand opportunities’ appropriate for a ‘flagship’ NHGRI program(s)?
2. What is NHGRI not doing that it should be doing?
3. How to balance ‘democratization’ of genome sequencing and benefits of consortia-based, large-scale pursuits?
4. How to properly tune a ‘flagship’ program’s(s’) funding level with its impact?
5. Should NHGRI develop a formal cost-sharing approach for large non-generic (e.g., disease-specific) projects?
6. How should NHGRI more efficiently obtain ‘commodity’ genome sequencing to meet programmatic needs?

Going Forward (Quickly)…

There are Likely Too Many Good Options

NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Advancing human health through genomics research