



genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Bioinformatics and Computational Biology in Genetics Research

House of Lords Inquiry on Genomic Medicine

Peter Good

Program Director, NHGRI

June 5, 2008





genome.gov

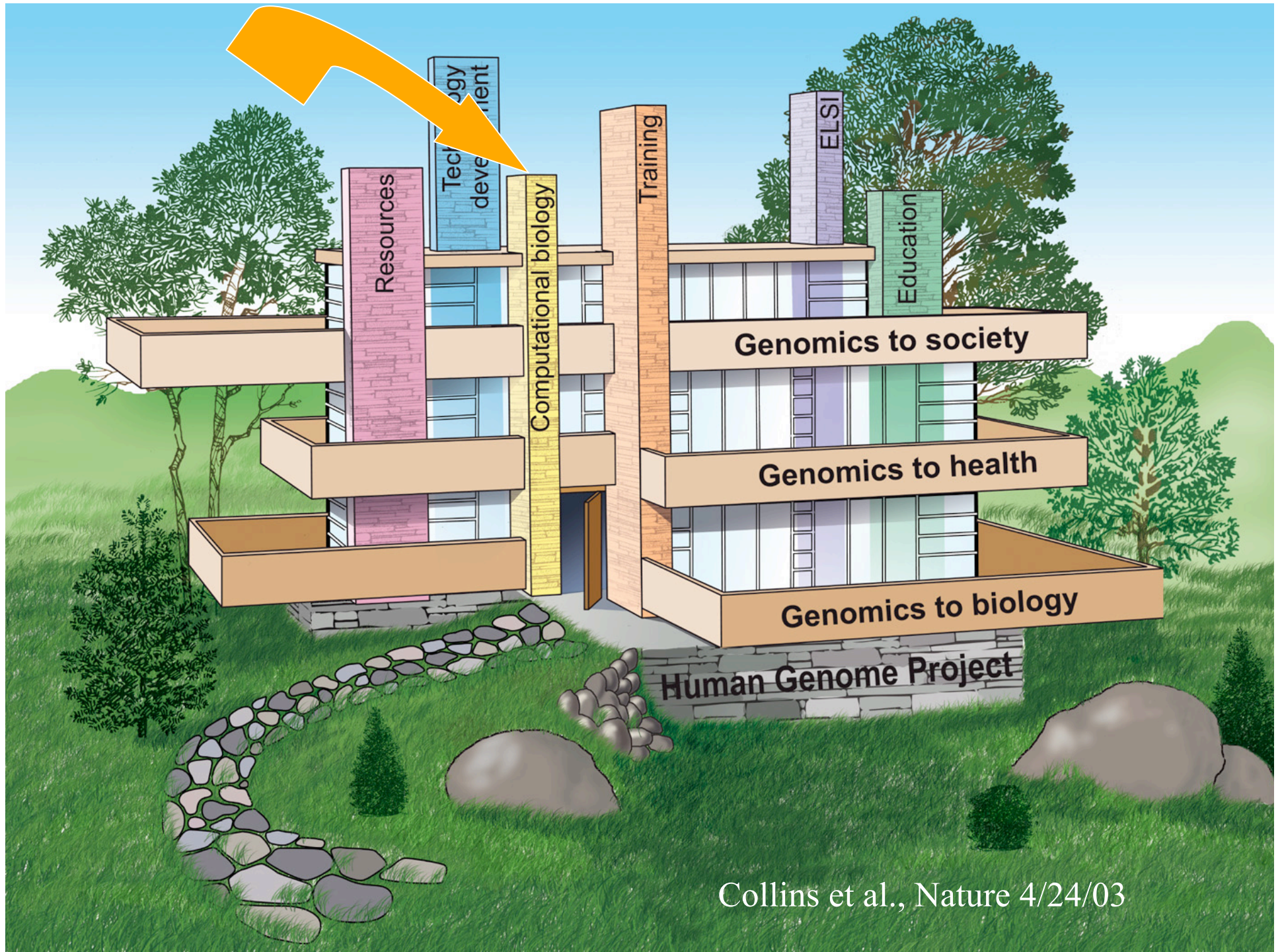
National Human Genome Research Institute

Advancing human health through genetic research

NHGRI Informatics Portfolio

- Resource Projects
- Computational research
- Resources for the human genome
- Challenges for the future





Collins et al., Nature 4/24/03



genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

NHGRI-Sponsored Programs

- **Role is integrative, creating scientific infrastructure**
- **Data types of wide-spread utility across wide areas of biomedical research**
- **Community Resource Projects**





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Importance of data sharing

- NHGRI supports comprehensive catalogues
- Pre-publication data release puts data into the hands of users (bioinformaticians)
- Ft. Lauderdale principles (Wellcome Trust/NHGRI)
 - Human Genome Project (and others) - reference sequence
 - HapMap - genetic variation
 - Mammalian Gene Collection - protein coding genes
 - Mouse Knock Out Mutant Project (KOMP) - mouse mutants
 - ENCODE - functional elements in the genome
 - GWAS - genetic factors that influence health and disease





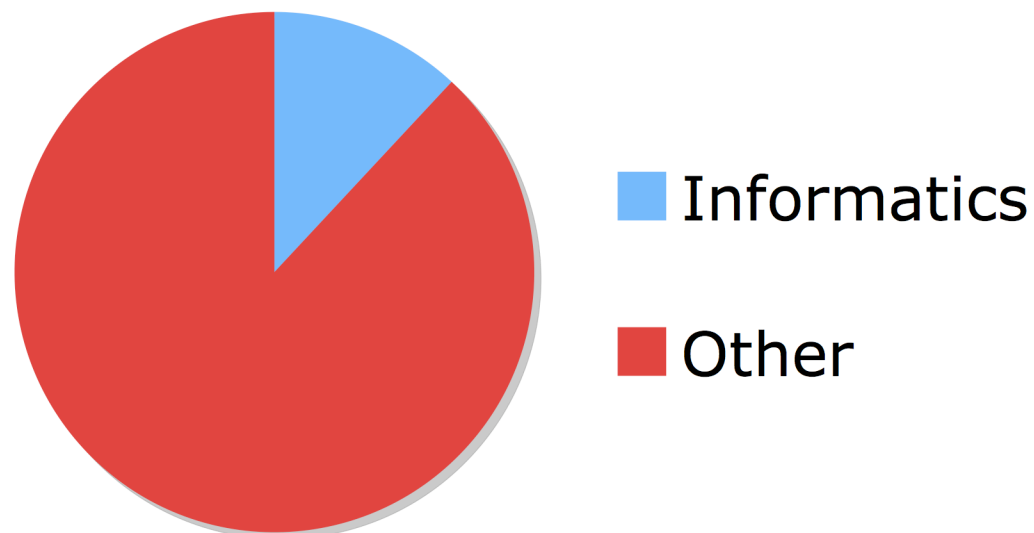
genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

NHGRI Informatics Budget

Extramural Informatics Funding



13.5 % of extramural funds for informatics





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

NHGRI Informatics

- Resource Projects
 - Model Organism Databases
 - Data Standards
 - Protein / Pathway Databases
- Technology Development (Research)
 - Extract information from genome datasets (sequence, microarray, phenotype, etc)





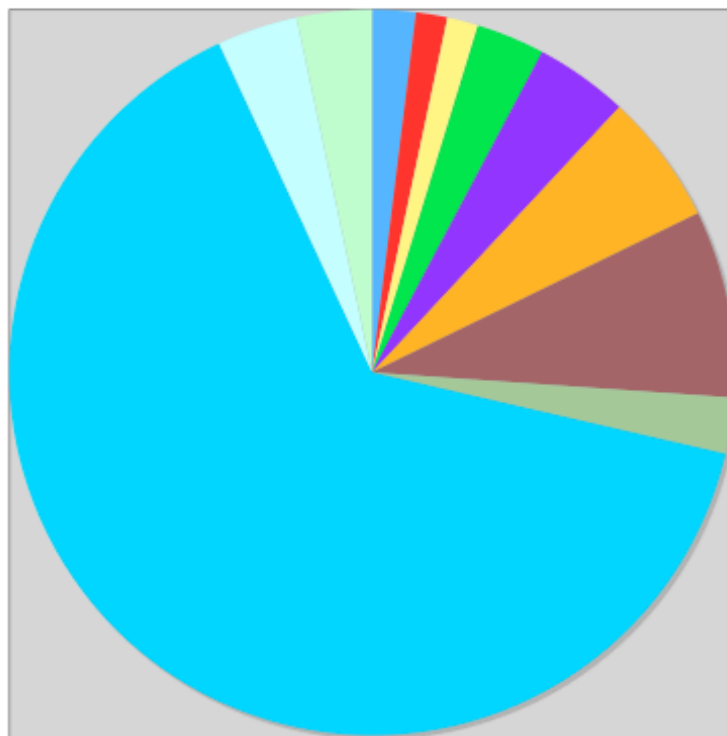
genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Informatics Funding pt 2

FY2007 Informatics Budget \$52,944,570



- Annotation
- Assembly
- Career
- DataAnalysis
- DatabaseTech
- DCC
- Elements
- GeneFinding
- GRIN
- Networks
- Software





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

MOD Functions

Link the biology to the sequence

- Provide a reference sequence and gene set to the community
- Location of genes and sequence features
- Locus page with gene information
 - What does protein do? What is the mutant phenotype? Where is it expressed?
- Guide to Literature
 - Ph.D.-level curators extract functional information
- Integration of other resources
 - Comparative genomics; large scale expression analyses; protein interactions; genetic interactions





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Demos

- WormBase www.wormbase.org
- Mouse Genome Informatics MGI





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Tools for interoperability

- Gene Ontology (GO) Consortium
 - Initial effort by MODs to share annotation expertise
 - Editorial office at EBI; all MODs participate
- Develop structured vocabularies (ontologies) to describe gene products
 - Biological process
 - What does it do? ie translational regulation
 - Molecular function
 - How does it work? ie RNA-binding protein
 - Cellular component
 - Where is it? ie cytoplasmic protein





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Generic Model Organism Database Project

- Develop generic tools that all MODs can use
- Promote a community of MOD software developers
- A goal of “off the shelf” components for new MOD
- Tools
 - Chado (database schema)
 - Generic Genome Browser (sequence viewer)
 - Textpresso (literature search)





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Other Informatics Resources

- UniProt: knowledgebase of protein sequence and function
 - Collaboration between EBI, Swiss Institute for Bioinformatics, and Georgetown University
 - Highly curated information about proteins
- Reactome: database of biological processes in humans
 - How do gene products interact?





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Pathway analysis of somatic mutations for TCGA

Interpreting mutation data tumors in enhanced by combing with information on biological pathways



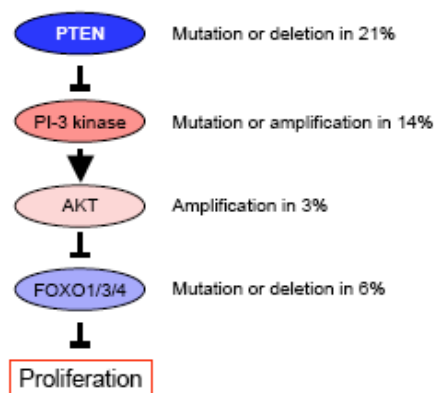


genome.gov

National Human Genome Research Institute

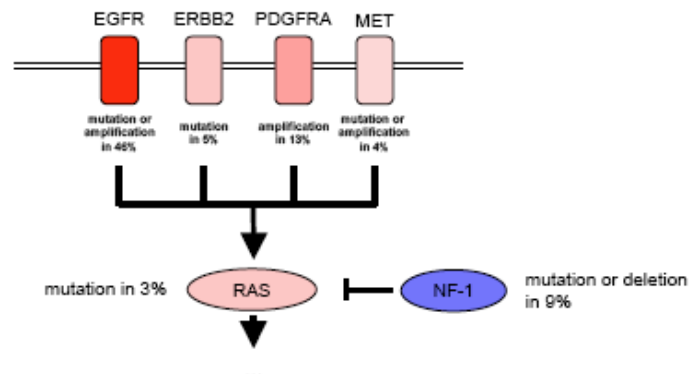
Advancing human health through genetic research

PI3-kinase / AKT signaling



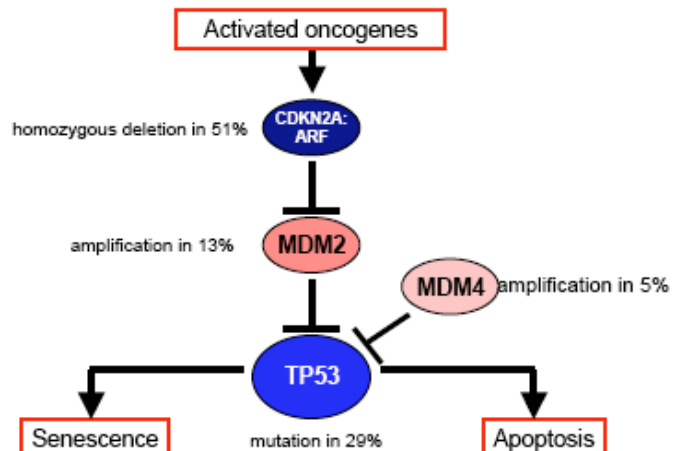
>30% of glioblastoma samples have DNA alterations in the PI3K-Akt pathway

RTK/Ras signaling



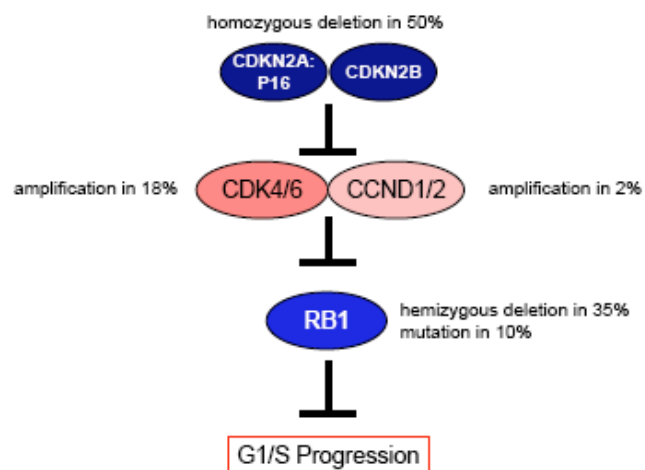
>60% of glioblastoma samples have DNA alterations in the RTK-Ras pathway

p53 signaling



~84% of glioblastoma samples have DNA alterations in the p53 pathway

RB signaling



~80% of glioblastoma samples have DNA alterations in the RB pathway





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

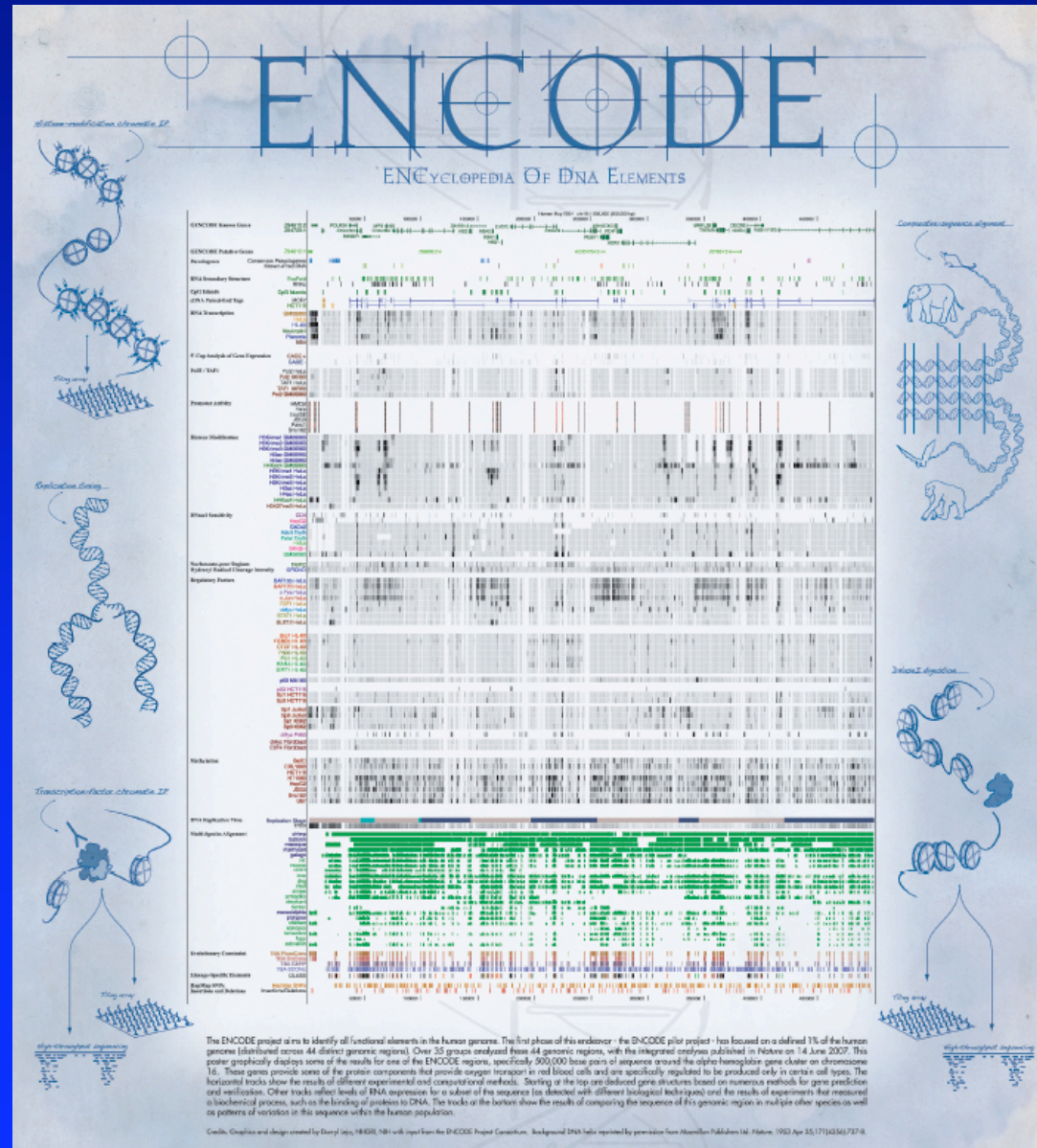
Informatics support for large projects

- HapMap, KOMP, ENCODE, 1000 Genomes, Human Microbiome, Epigenomics
- All have Data Coordinating Centers funded as part of the project
- Example: ENCODE Project at the UCSC Genome Browser





Example of data at the ENCODE Data Coordination Center





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Technology Development

- Largely unsolicited grants - support creative researchers
- Examples
 - Computational methods to find functional elements (genes, promoters, enhancers, etc)
 - Understand gene regulation networks
 - Computational predictions of protein function





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Resources for the Human Genome

- Genome Browsers
 - UCSC (NHGRI)
 - Ensembl (EBI)
 - MapViewer (NCBI/NLM)
- Provide different views of gene features





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Resources for the Human Genome

- OMIM: Online Mendelian Inheritance in Man
 - Funded by NHGRI/NLM
 - Description of mutations and clinical phenotypes for genetic disorders
- GeneTests / GeneReviews
 - Funded by NHGRI/NLM
 - Medical genetics information resource developed for physicians, other healthcare providers, and researchers
 - Online publication of expert-authored disease reviews
Genetic Disease Online Reviews Index





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

NIH Roadmap

- Support cross-cutting, trans-NIH programs
 - Identify gaps in NIH portfolio
- Bioinformatics and Computational Biology identified a key area
- Solicitation for National Centers for Biomedical Computing to generate computational infrastructure (software tools)





NCBC Centers

Informatics for Integrating Biology and the Bedside
Brigham and Women's Hospital

National Alliance for Medical
Imaging Computing

National Center for
Integrative Biomedical Informatics
University of Michigan

National Center for Biomedical Ontology
Physics-based Simulation of Biological Structures
Stanford University

Center for Computational Biology
University of California at Los Angeles

National Center for Multi-Scale
Study of Cellular Networks
Columbia University



Breaking Events: [NCBC Biositemaps Page](#)



genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Challenges

- More and more data



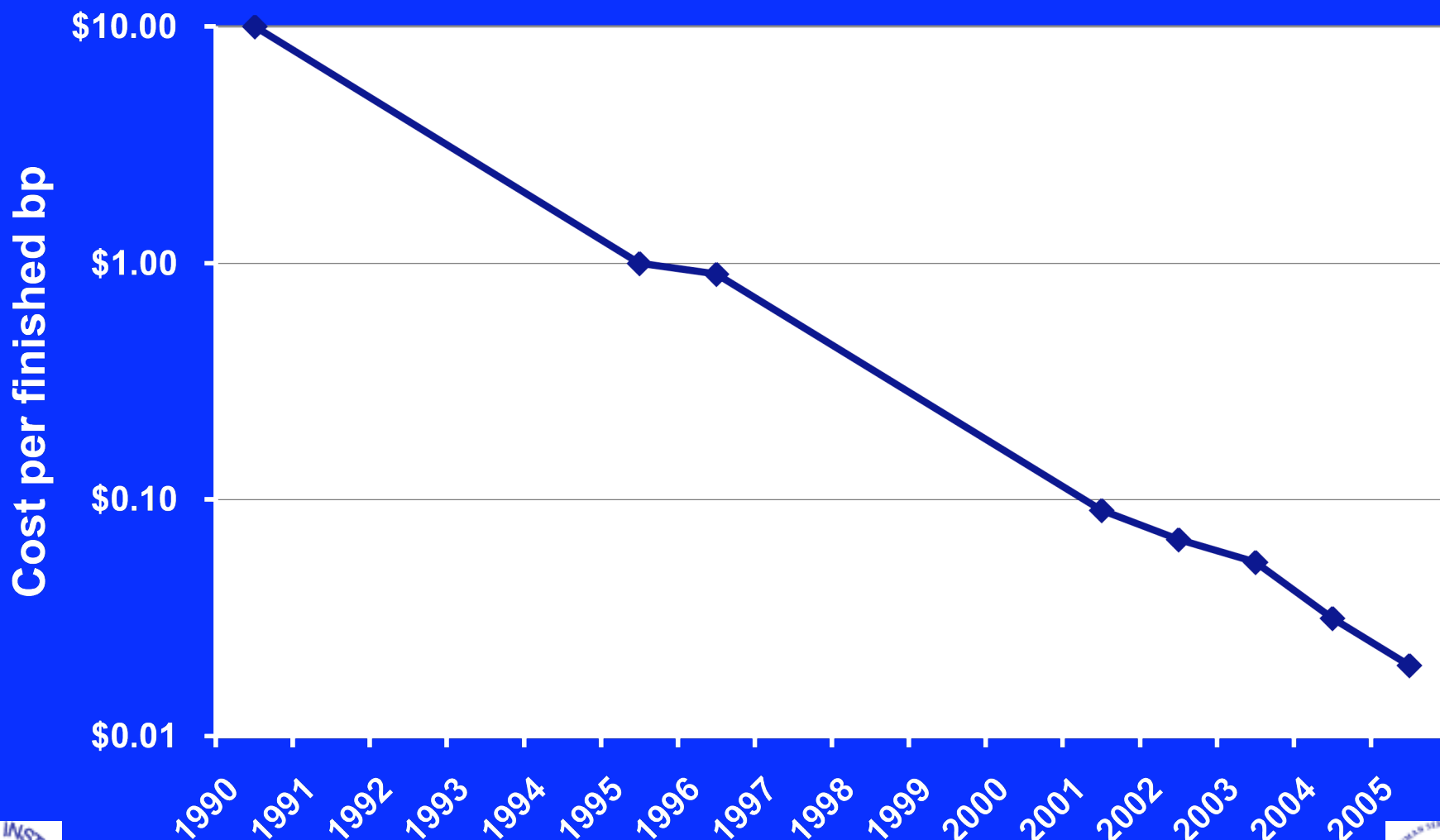


genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Decrease in the Cost of Finished DNA Sequencing





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Challenges

- More and more data
- New technologies, new data analysis methods (and scale) are needed





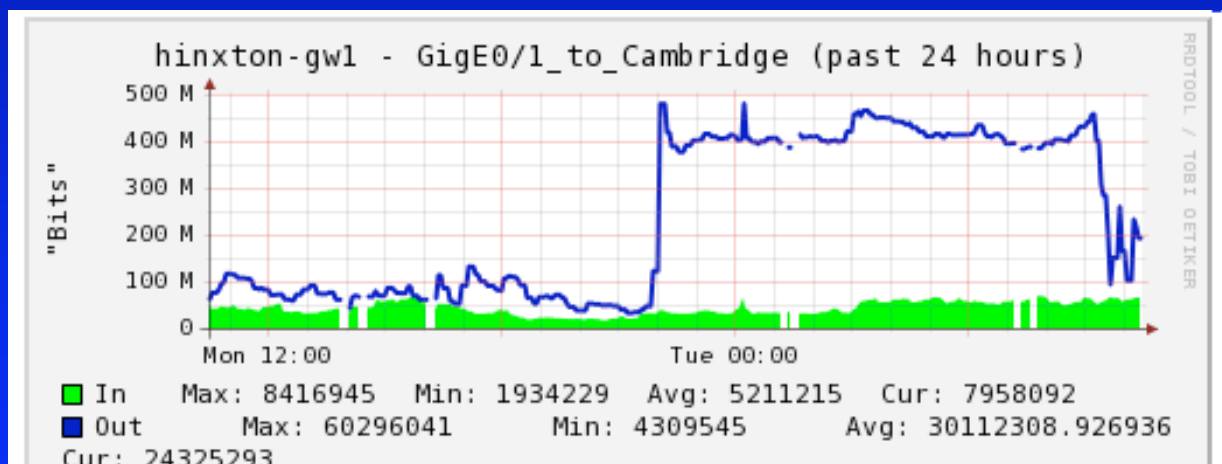
genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Challenges in “drinking from the firehose”

- Data handling, informatics resources: a LOT of data—the *initial* deposition increased the total sequence data available in the public domain by 10%, overnight



- Analysis, analysis, analysis...
- Samples, with appropriate consent for use in genomic studies and *data release*





genome.gov

National Human Genome Research Institute

Advancing human health through genetic research

Challenges

- More and more data
- New technologies, new data analysis methods (and scale) are needed
- Funding for resources
 - Database vs Science?
- Recognition of computational biologists
- Integration of “wet” and “dry” researchers
- Training

