Looking Back, Looking Forward: It's Still Not the Post-Genomic Era

Francis S. Collins, M.D., Ph.D.
Director, National Institutes of Health

10th Anniversary of the Completion of the Human Genome Project

April 25, 2013

April, 1953 → April, 2003

No. 4356 April 25, 1953 NATURE

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

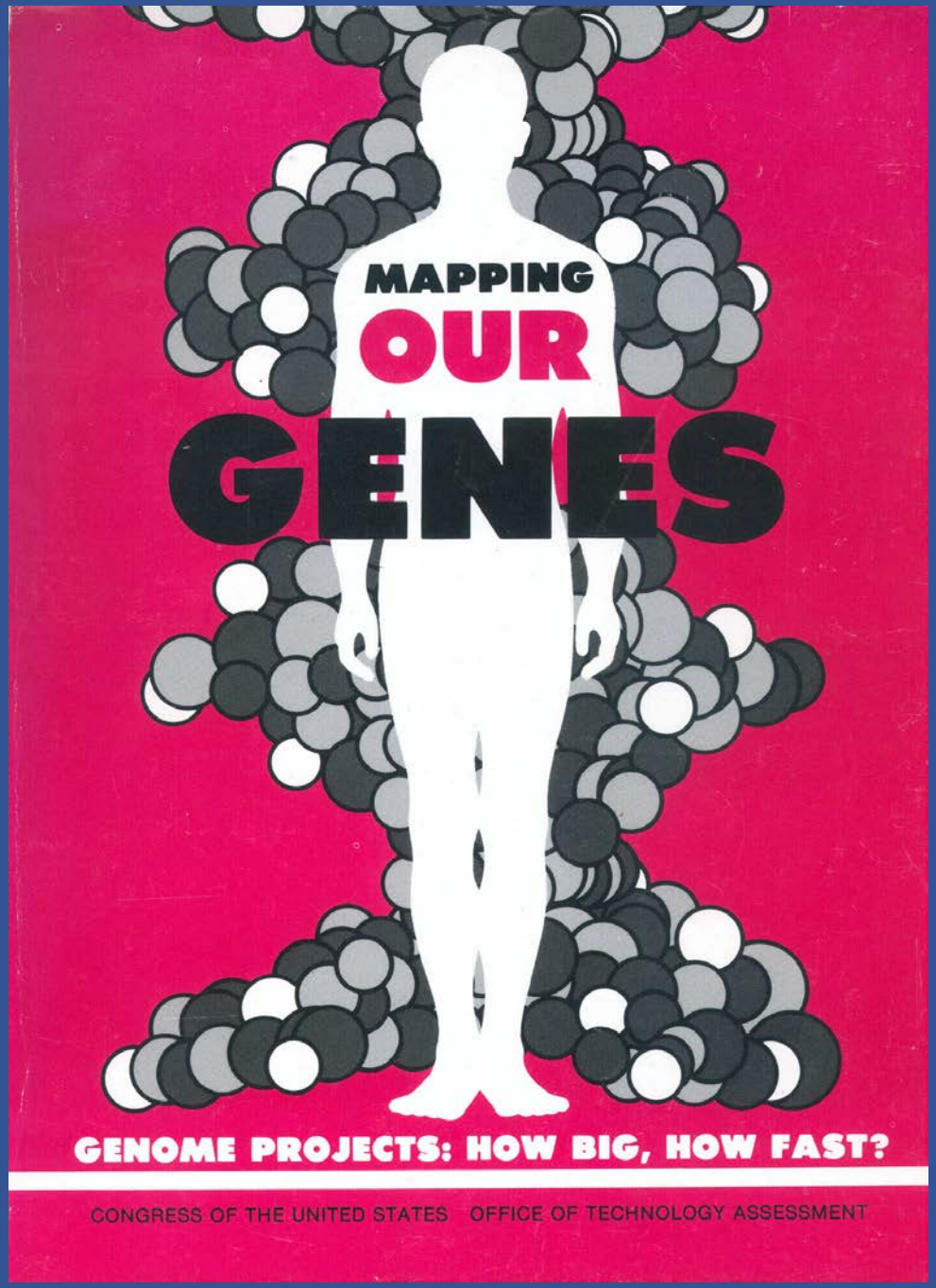
A Structure for Deoxyribose Nucleic Acid

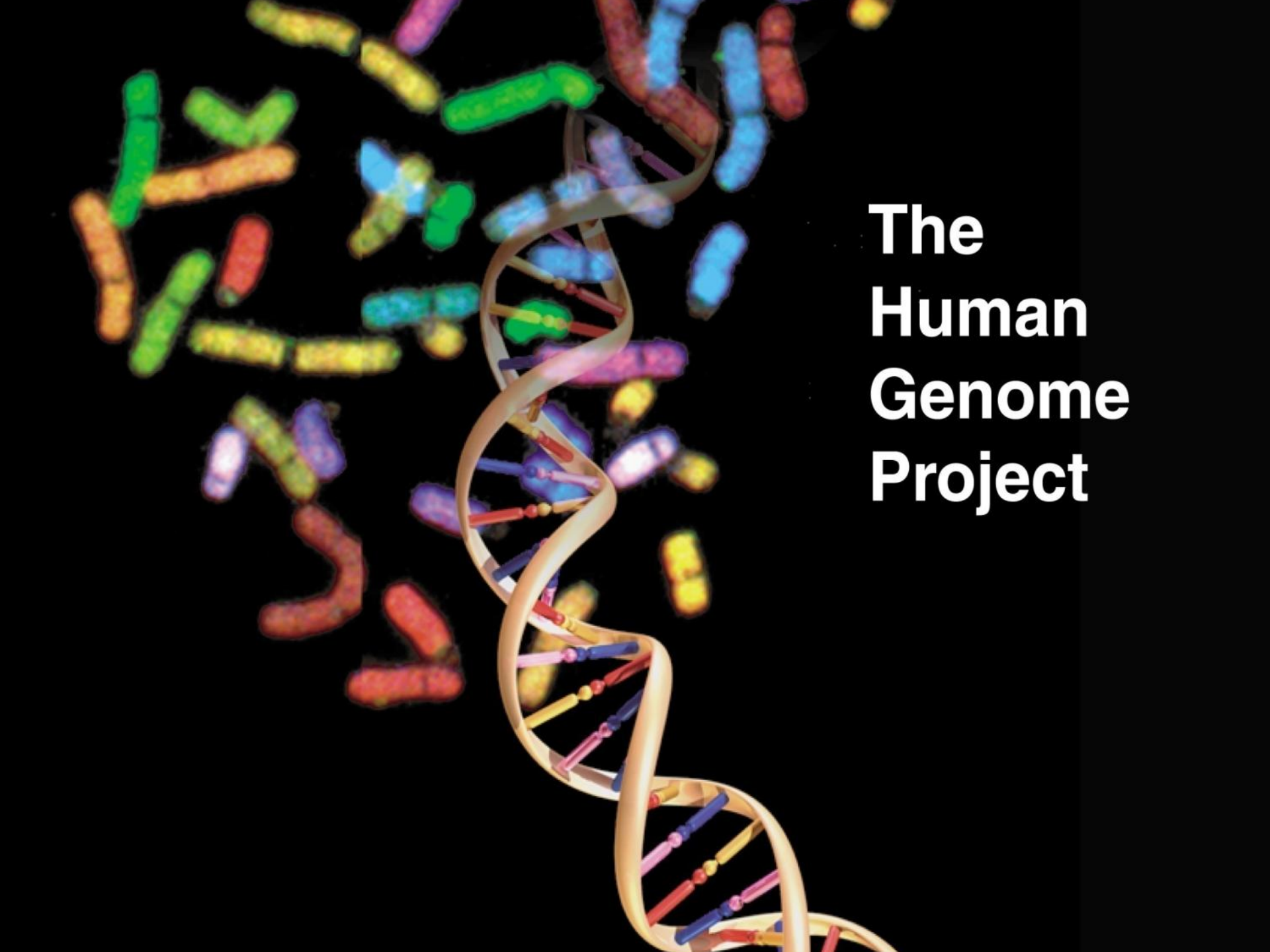
J. D. Watson
F. H. C. Crick

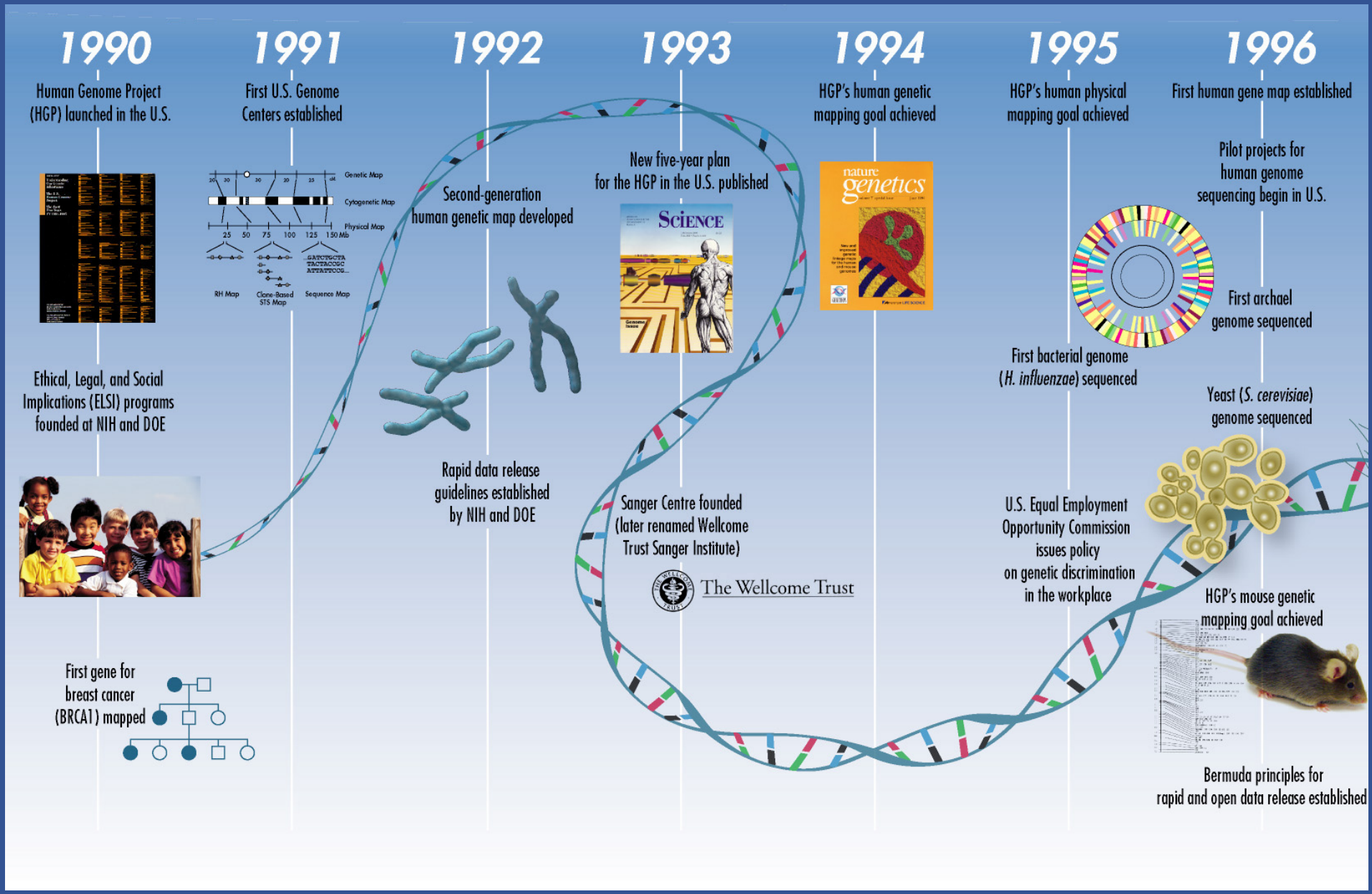
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

MAPPING AND SEQUENCING THE HUMAN GENOME

NATIONAL RESEARCH COUNCIL

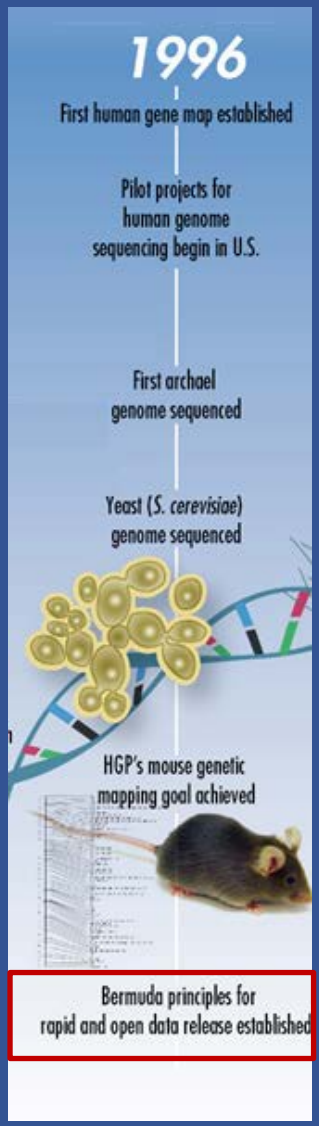


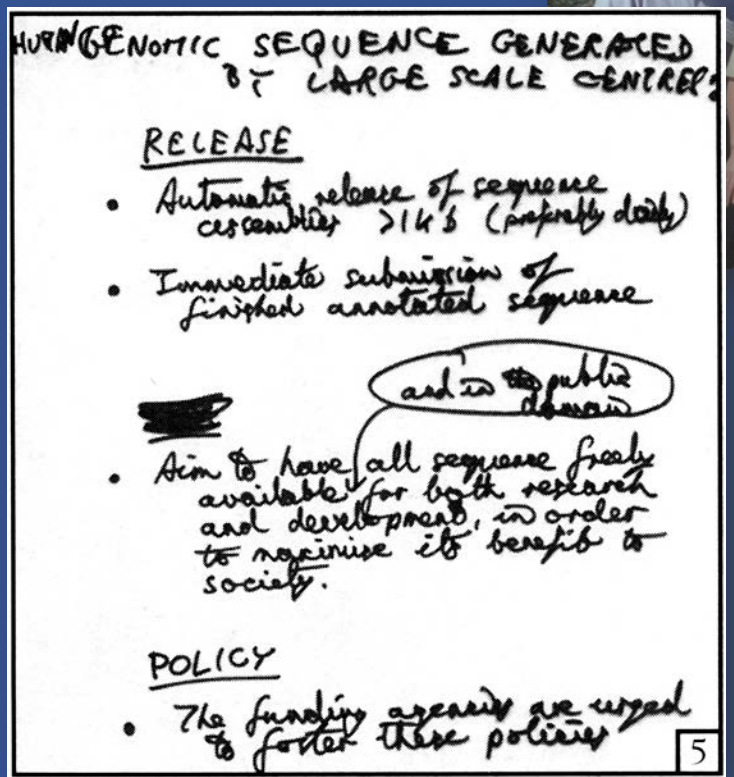

The
Human
Genome
Project



Collins et al., Nature 4/24/03

Laying the Foundation for Open Access: Bermuda, 1996




HUMAN GENOMIC SEQUENCE GENERATED BY LARGE SCALE CENTRES

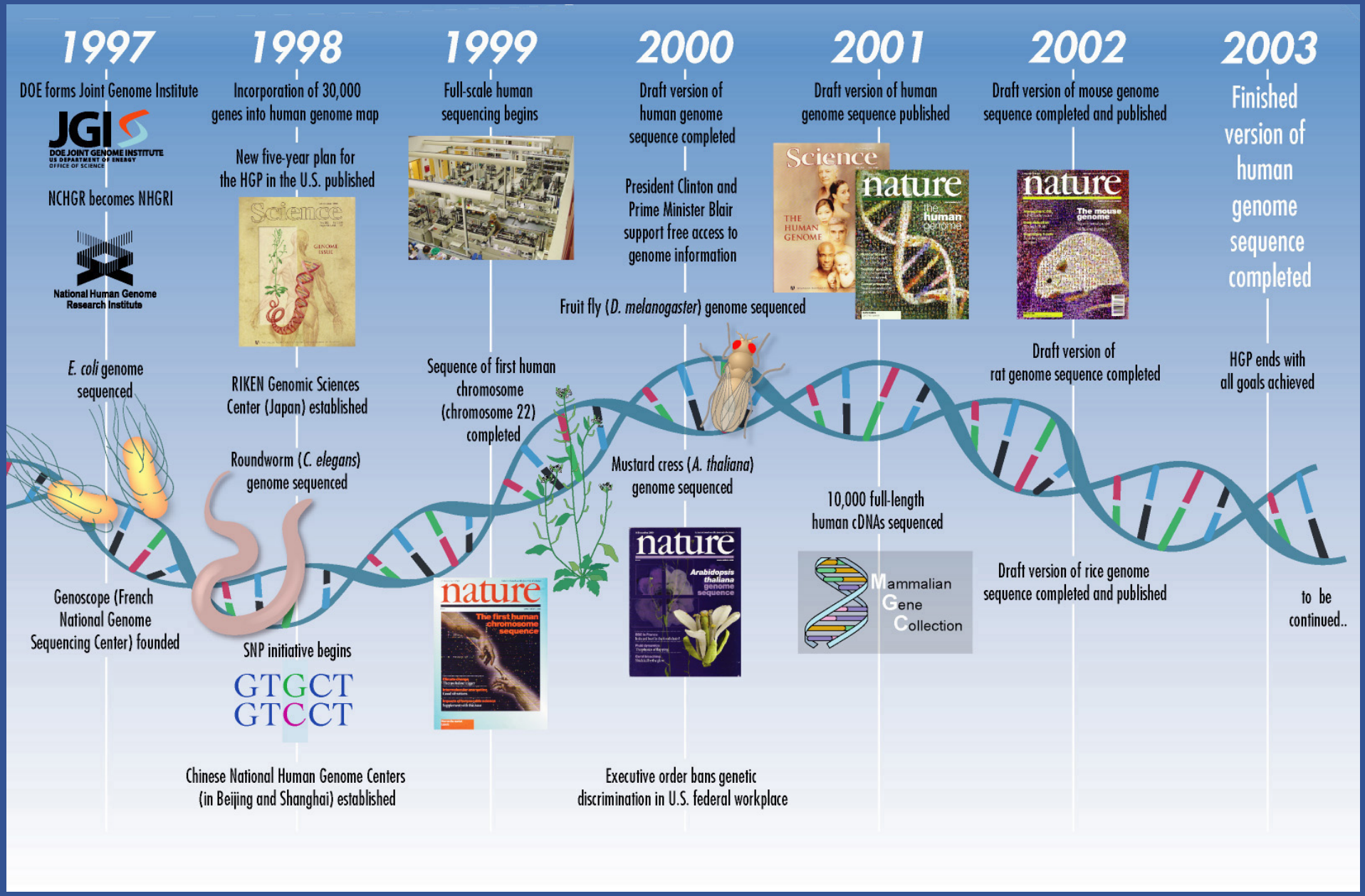
RELEASE

- Automatic release of sequence assemblies >1 kb (preferably daily)
- Immediate submission of finished annotated sequence (and in the public domain)
- Aim to have all sequence freely available for both research and development, in order to maximise its benefit to society.

POLICY

- The funding agencies are urged to foster these policies

5



Collins et al., Nature 4/24/03

2000

Draft version of human genome sequence completed

President Clinton and Prime Minister Blair support free access to genome information

THE WHITE HOUSE

Office of the Press Secretary

For Immediate Release June 25, 2000

PRESIDENT CLINTON ANNOUNCES THE COMPLETION OF THE FIRST SURVEY OF THE ENTIRE HUMAN GENOME

Hails Public and Private Efforts Leading to This Historic Achievement

June 26, 2000

Today, at a historic White House event with British Prime Minister Tony Blair, President Clinton announced that the international Human Genome Project and Celera Genomics Corporation have both completed an initial ...int for human beings. ...ublic and private

theguardian

News › Science › Genetics

Scientists finish first draft of DNA blueprint

Interactive guide

Tim Radford, science editor
The Guardian, Monday 26 June 2000 03.10 EDT

Scientists in London and Washington will announce today that they have completed the "first draft" of the entire blueprint of human life, described as the most important scientific effort humankind has ever mounted, including splitting the atom and going to the moon.

nature International weekly journal of science

News

Nature **405**, 983-984 (29 June 2000) | doi:10.1038/35016696

World leaders heap praise on human genome landmark

Colin Macilwain

NATIONAL INSTITUTES OF HEALTH

NIH NEWS RELEASE

NATIONAL INSTITUTES OF HEALTH — National Human Genome Research Institute

U.S. Department of Engery

FOR IMMEDIATE RELEASE
Monday, June 26, 2000
10:30 a.m. EST

Contact:
Cathy Yarbrough, NHGRI
(301) 594-0954

International Human Genome Sequencing Consortium Announces "Working Draft" of Human Genome

Science

The New York Times ON THE WEB

June 27, 2000

Genetic Code of Human Life Is Cracked by Scientists

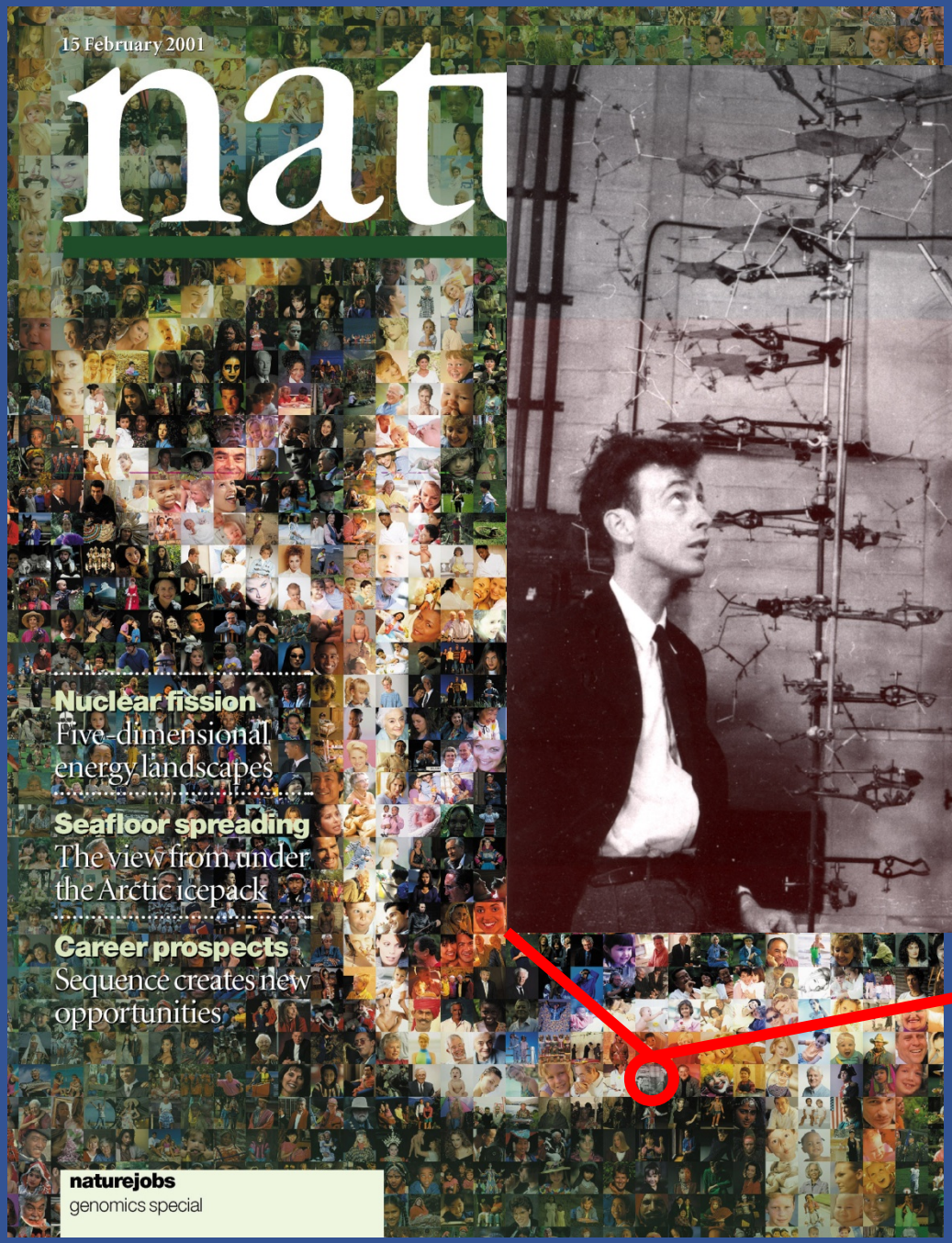
By NICHOLAS WADE

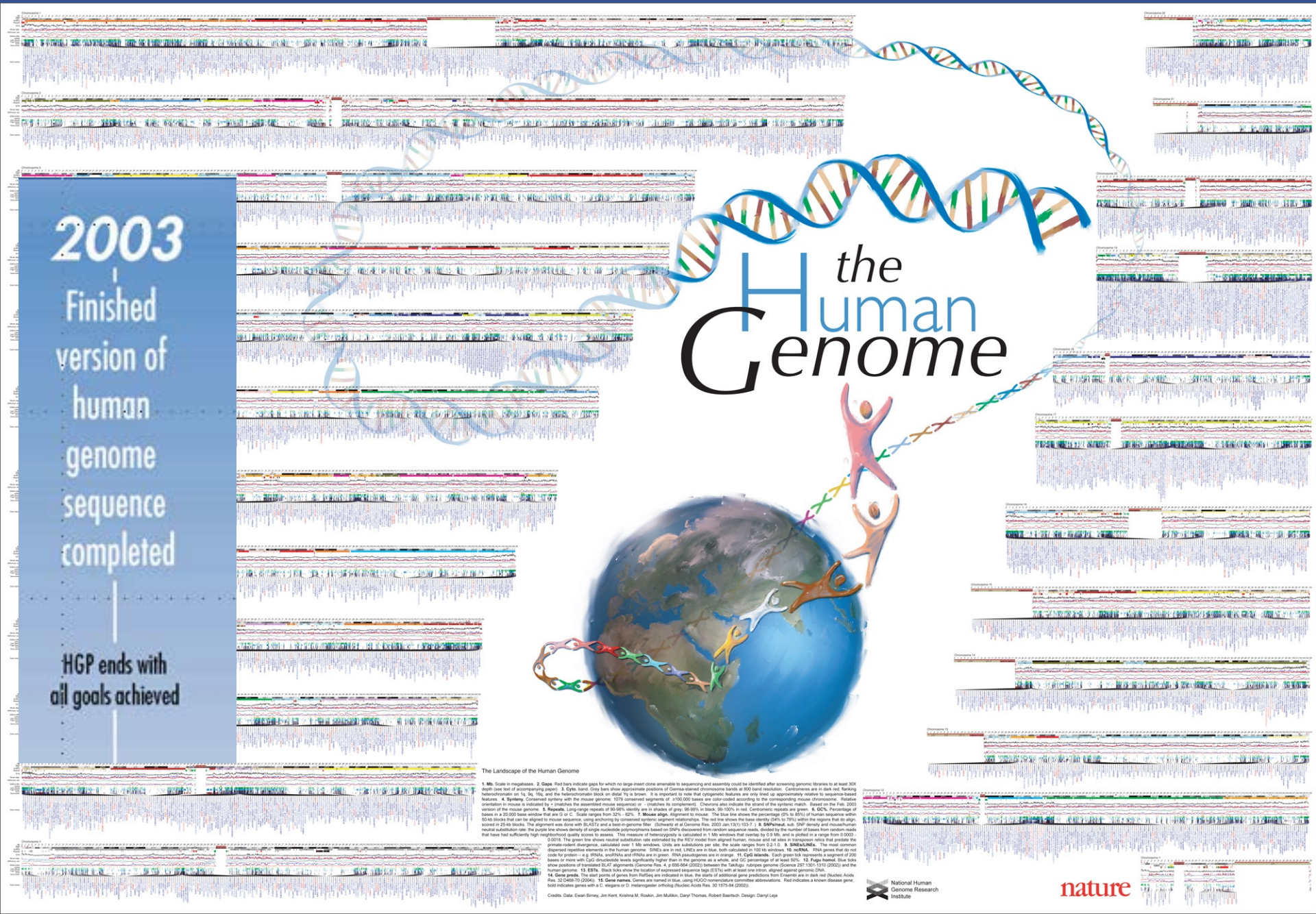
WASHINGTON, June 26 -- In an achievement that represents a pinnacle of human self-knowledge, two rival groups of scientists said today that they had deciphered the hereditary script, the set of instructions that defines the human organism.

Reuters

2001

Draft version of human genome sequence published





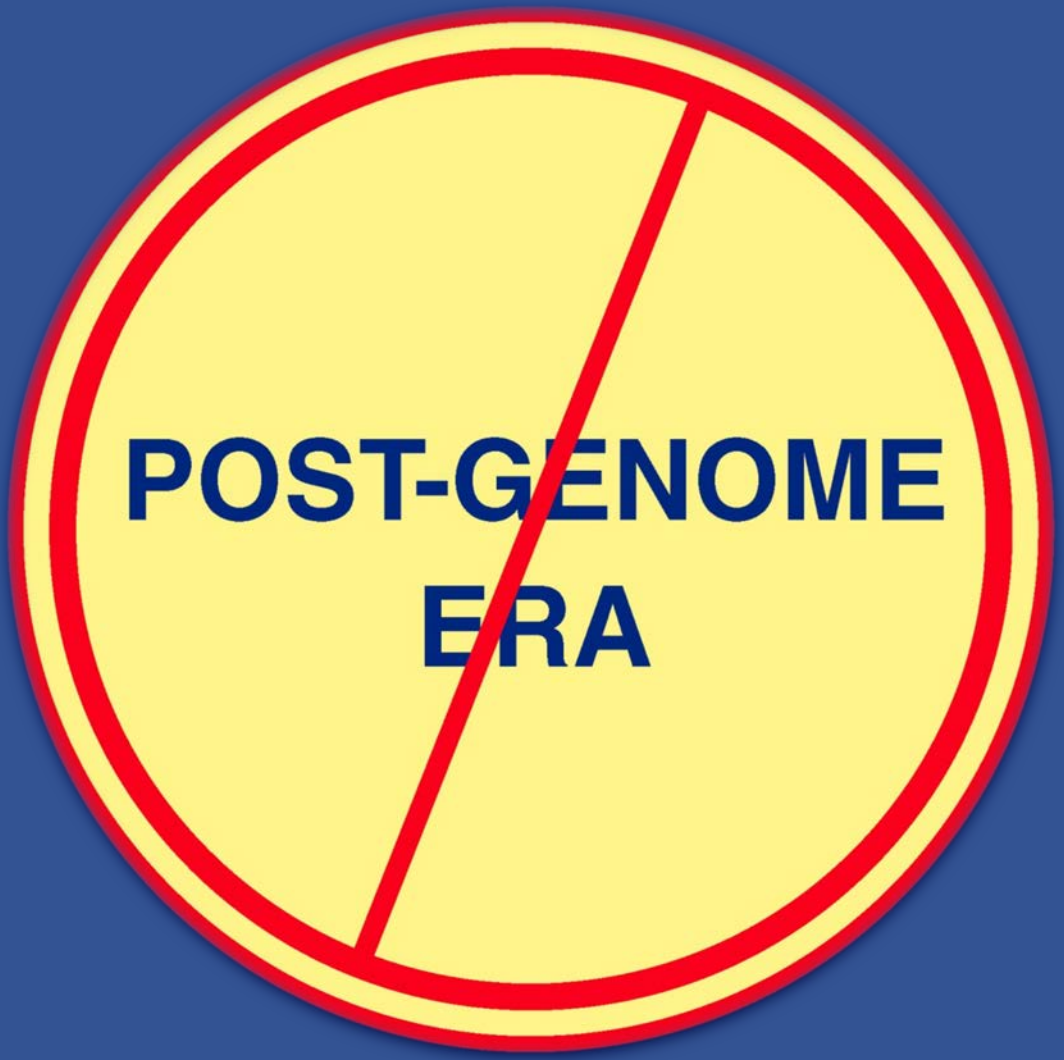
The Landscape of the Human Genome

1. Mb. Scale in megabases. **2. Gaps.** Red bars indicate gaps for which no large-insert clone amenable to sequencing and assembly could be identified after screening genomic libraries to at least 30X depth (see text of accompanying paper). **3. Cyto. band.** Grey bars show approximate positions of Giemsa-stained chromosome bands at 800 band resolution. Centromeres are in dark red; flanking heterochromatin on 1q, 9q, 16q, and the heterochromatin block on distal 19q is brown. It is important to note that cytogenetic features are only lined up approximately relative to sequence-based features. **4. Synteny.** Conserved synteny with the mouse genome: 1079 conserved segments of ≥100,000 bases are color-coded according to the corresponding mouse chromosome. Relative orientation in mouse is indicated by + (matches the assembled mouse sequence) or - (matches its complement). Chevrons also indicate the strand of the syntenic match. Based on the Feb. 2003 version of the mouse genome. **5. Repeats.** Long-range repeats of 90-98% identity are in shades of grey; 98-99% in black; 99-100% in red. Centromeric repeats are green. **6. GC%.** Percentage of bases in a 20,000 base window that are G or C. Scale ranges from 32% - 62%. **7. Mouse align.** Alignment to mouse. The blue line shows the percentage (0% to 85%) of human sequence within 50-kb blocks that can be aligned to mouse sequence, using anchoring by conserved syntenic segment relationships. The red line shows the base identity (94% to 79%) within the regions that do align, scored in 25-kb blocks. The alignment was done with BLASTz and a best-in-genome filter. (Schwartz et al Genome Res. 2003 Jan;13(1):103-7.) **8. SNPsinout.** sub. SNP density and mouse/human neutral substitution rate: the purple line shows density of single nucleotide polymorphisms based on SNPs discovered from random sequence reads, divided by the number of bases from random reads that have had sufficiently high neighborhood quality scores to assess. This measure of heterozygosity is calculated in 1 Mb windows that overlap by 0.9 Mb, and is plotted in a range from 0.0003 - 0.0018. The green line shows neutral substitution rate estimated by the REV model from aligned human, mouse and rat sites in transposon relics that predate the primate-rodent divergence, calculated over 1 Mb windows. Units are substitutions per site; the scale ranges from 0.2-1.0. **9. SINEs/LINEs.** The most common dispersed repetitive elements in the human genome. SINEs are in red, LINEs are in blue, both calculated in 100 kb windows. **10. ncRNA.** RNA genes that do not code for protein – e.g. tRNAs, snoRNAs and rRNAs are in green. RNA pseudogenes are in orange. **11. CpG islands.** Each green tick represents a segment of 200 bases or more with CpG dinucleotide levels significantly higher than in the genome as a whole, and GC percentage of at least 50%. **12. Fugu homol.** Blue ticks show positions of translated BLAT alignments (Genome Res. 4, p 656-664 (2002)) between the Takifugu rubripes genome (Science 297:1301-1310 (2002)) and the human genome. **13. ESTs.** Black ticks show the location of expressed sequence tags (ESTs) with at least one intron, aligned against genomic DNA. **14. Gene preds.** The start points of genes from RefSeq are indicated in blue, the starts of additional gene predictions from Ensembl are in dark red (Nucleic Acids Res. 32:D468-70 (2004)). **15. Gene names.** Genes are named in blue, using HUGO nomenclature committee abbreviations. Red indicates a known disease gene, bold indicates genes with a C. elegans or D. melanogaster ortholog (Nucleic Acids Res. 30:1575-84 (2002)).

Credits: Data: Ewan Birney, Jim Kent, Krishna M. Roskin, Jim Mullikin, Daryl Thomas, Robert Baertsch Design: Darryl Leja

National Human Genome Research Institute

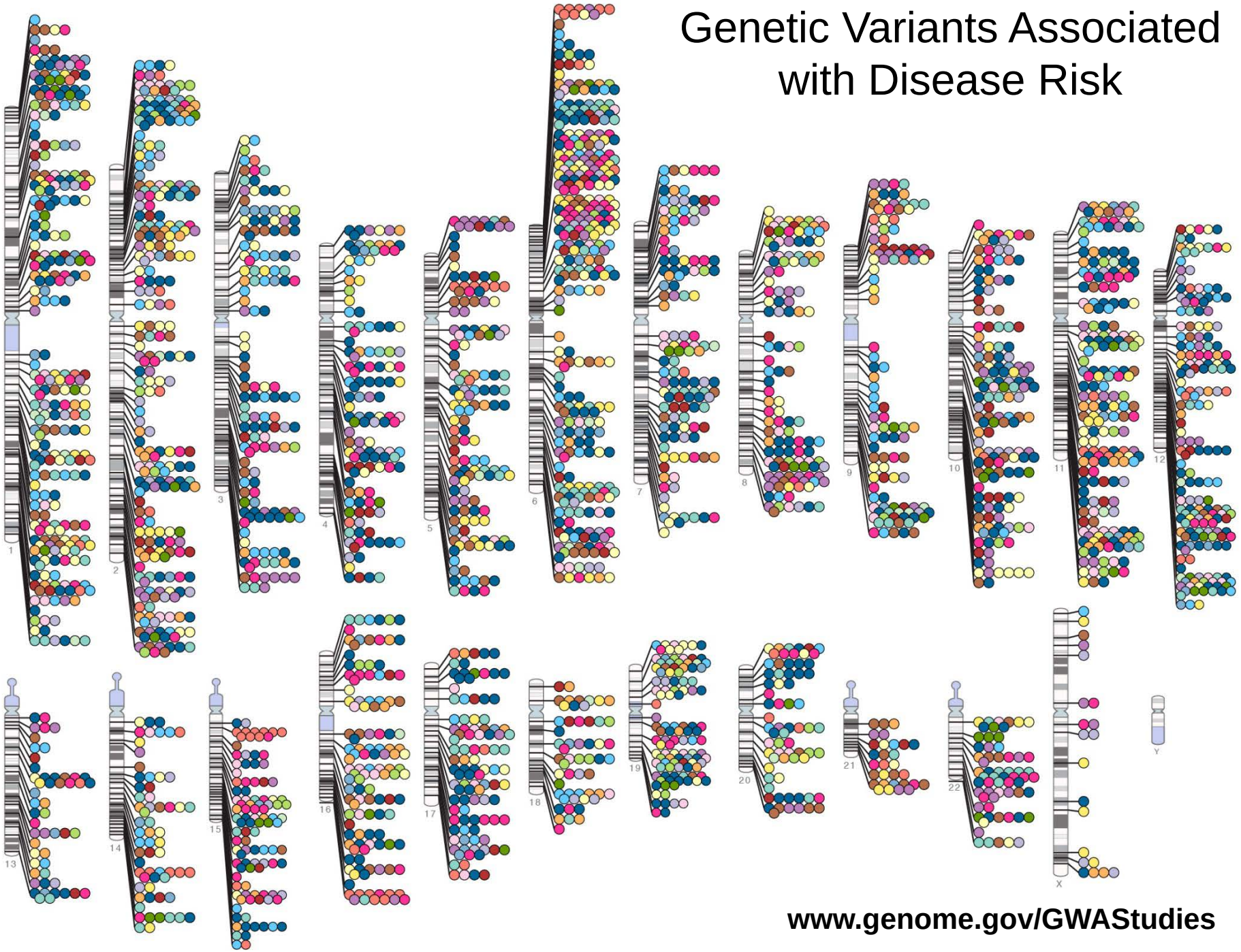
nature


POST-GENOME
ERA

The HapMap Project

- Produced genotypes for 3.5 million SNPs in 270 samples
- Data show allele frequencies, haplotypes, and associations among the SNPs
- Data were used to develop SNP chips for genome wide association (GWAS) studies
- Chips allow GWAS studies to find gene regions containing variants that affect health and disease

Genetic Variants Associated with Disease Risk



www.genome.gov/GWAStudies

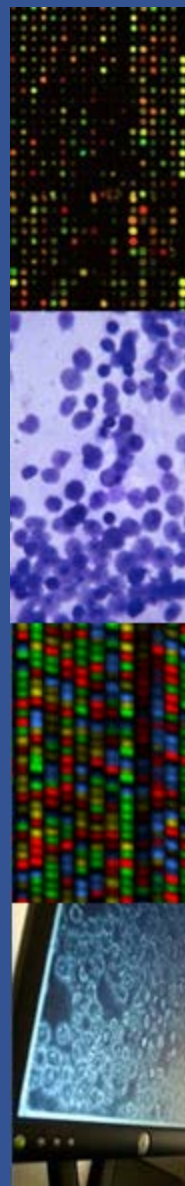
1000 Genomes in the Cloud

- 1000 Genomes Project (www.1000genomes.org)
  - Leverages recent improvements in next generation sequencing technology
  - Largest set of data on human genetic variation
- Data now freely available in Amazon's cloud
  - 200 terabytes of data
  - Researchers who would not otherwise have capacity or computing systems now have access to this data

Low coverage and exome sequencing now complete on 2500 samples!

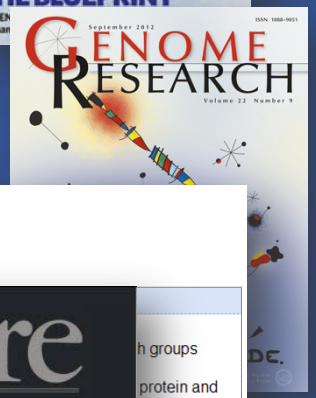
The Cancer Genome Atlas (TCGA) and the International Cancer Genomics Consortium

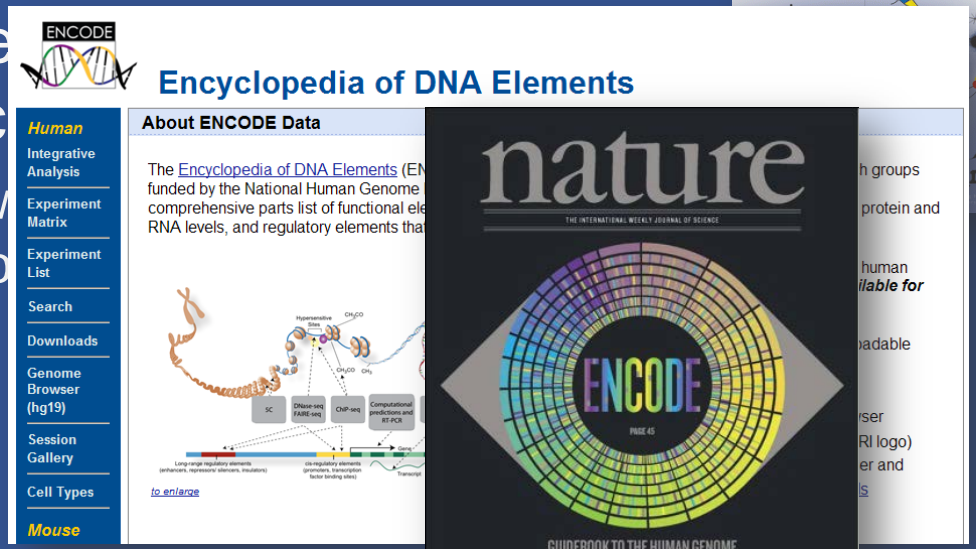
- Comprehensive, collaborative effort to map genomic changes in major types, subtypes of cancer
- NIH pilot (2006) expanded to 25 tumor types (2009)
  - Every case (tumor + germline) gets comprehensive characterization (SNP, exome, mi/RNAseq, methylation)
  - All data available pre-publication once quality controlled
  - 6,600+ cases now in pipeline
    - Seek to complete 11,000 by end of 2014
- International Cancer Genomics Consortium further expands program to many other tumor types
- **Ultimate goal:** improve our ability to diagnose, treat, and prevent cancer



ENCODE:
Encyclopedia of DNA Elements

- Launched in 2003, to develop comprehensive catalog of functional elements in the human genome
  - Related model organism projects
- Community resource project
  - All data free and available for immediate use
- Update 2012: 30+ pape
  findings, innovative acc
  - First global, detailed viev
    elements in human geno





ENCODE

www.encodeproject.org

nature ENCODE explorer

THREADS

PAPERS

01 02 03 04 05 06 07 08 09 10 11 12 13

THREAD OVERVIEW

Epigenetic regulation of RNA processing

ENCODE data explores properties of RNA processing and its relationship to histone modifications

Read Thread

What is ENCODE? | Threads: a new approach | Guide to the ENCODE explorer

“Big Data”: Challenges for Biomedicine in an Era of Massive Data Sets

Recent explosion of biomedical data

- Challenge: find ways to optimize data that
  - Speed discovery and innovation
  - Improve nation’s health, economy
- NIH responds to the challenge
  - New internal governing/oversight bodies
  - New trans-NIH initiative: Big Data to Knowledge (BD2K)
  - New leadership position: Associate Director for Data Science

Evelyn Hockstein for The New York Times

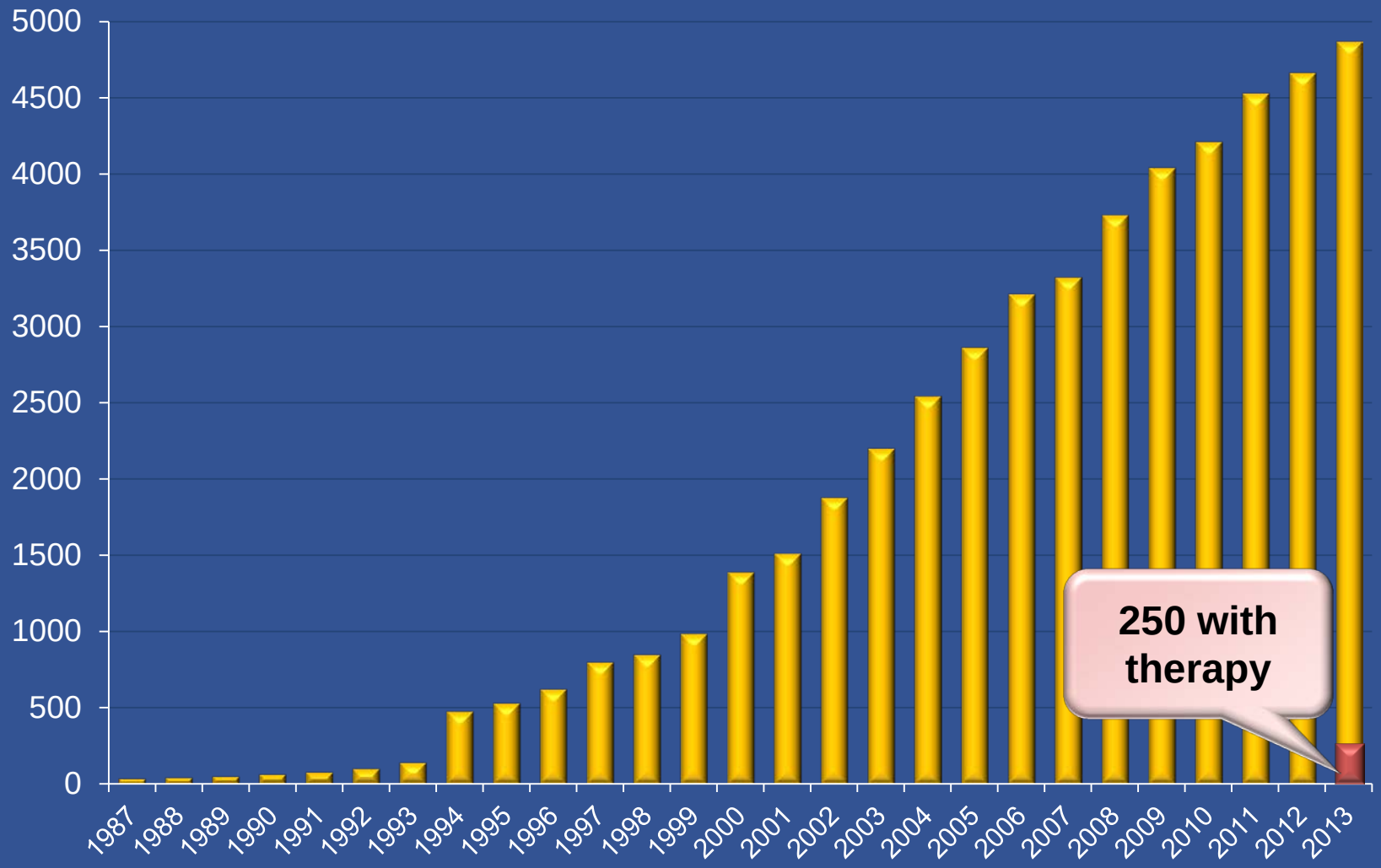
Eric
Director for Data Science

Disorders with Known Molecular Basis

Source: Online *Mendelian Inheritance in Man*, Morbid Anatomy of the Human Genome


A
CELEBRATION

Disorders with Known Molecular Basis



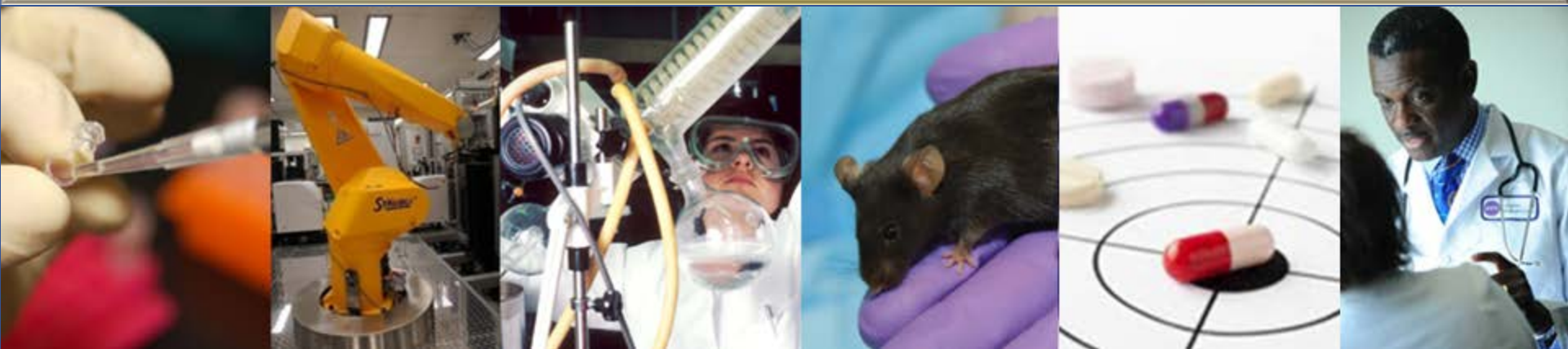
Source: Online *Mendelian Inheritance in Man*, Morbid Anatomy of the Human Genome

National Center for Advancing Translational Sciences (NCATS)

Mission:

To catalyze the generation of innovative methods and technologies that will enhance the development, testing, and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.

http://ncats.nih.gov/



the WHITE HOUSE

PRESIDENT OBAMA IS CALLING ON THE SCIENCE COMMUNITY
TO JOIN HIM IN PURSUING A GRAND CHALLENGE

BRAIN INITIATIVE

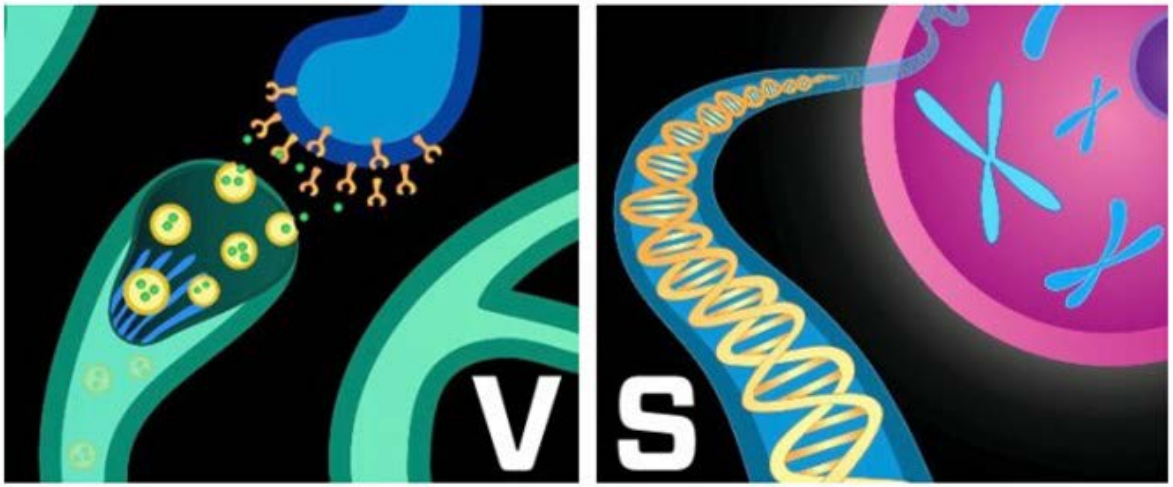
BRAIN RESEARCH
THROUGH ADVANCING
INNOVATIVE
NEUROTECHNOLOGIES

WIRED

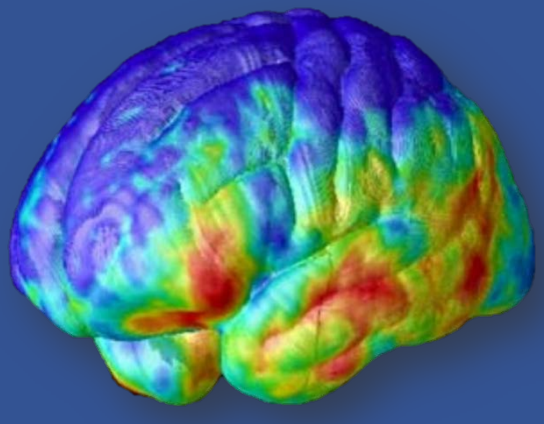
Now or Then: Which Big Science Project Are These Scientists Worked Up About?

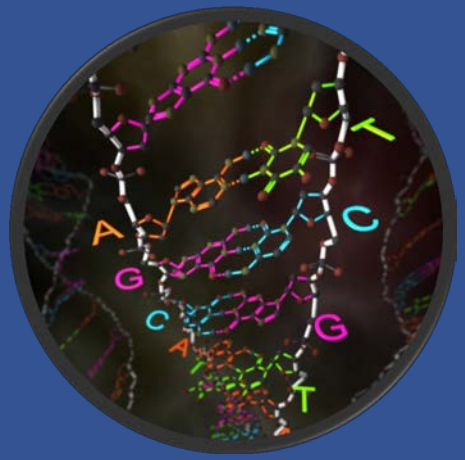
BY GREG MILLER 04.05.13 6:30 AM

Follow @dosmonos

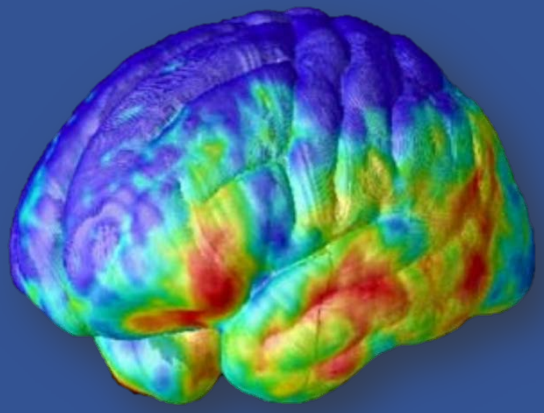


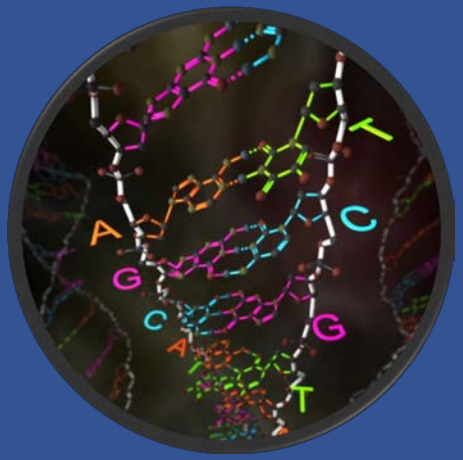
Now *or* Then





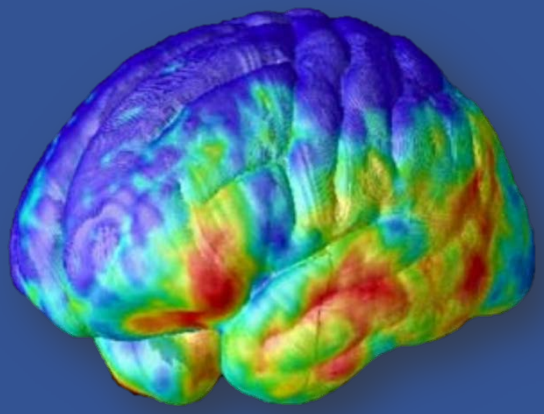
Now ***or*** **Then**

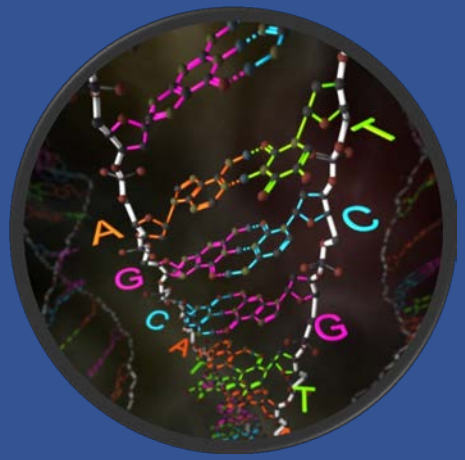




The ? project is bad science, it's unthought-out science, it's hyped science.

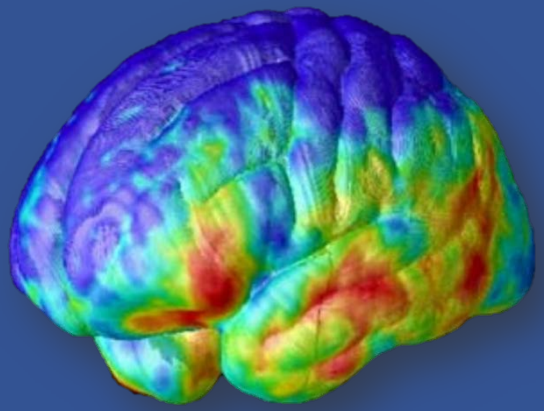
Now ***or*** **Then**

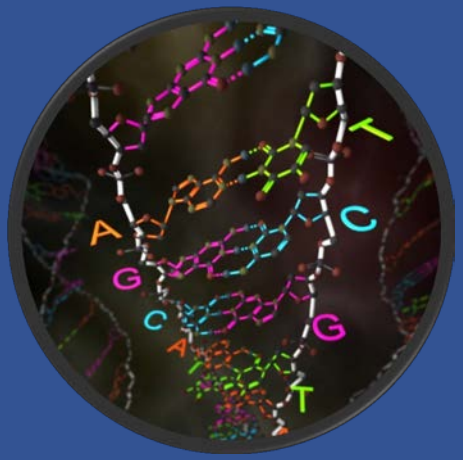




I believe the scientific paradigm underlying this ? project is, at best, out of date and at worst, simply wrong.

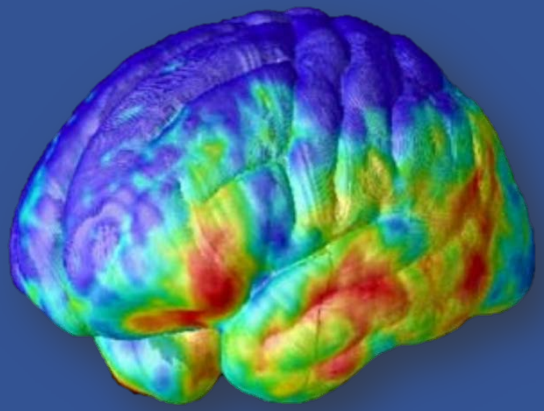
Now ***or*** **Then**

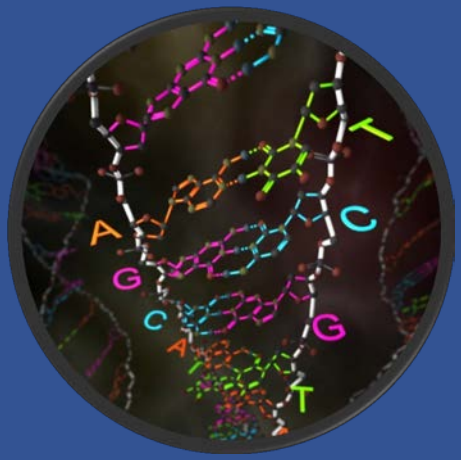




Concentrating hundreds of millions of dollars on this one megaproject in the era of ? budget cuts is sure to starve hundreds of small, more promising biomedical research projects.

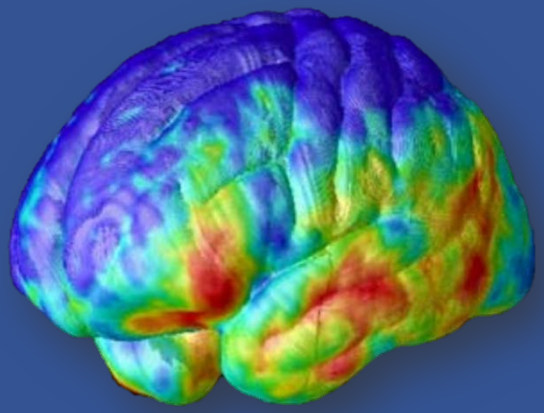
Now ***or*** **Then**

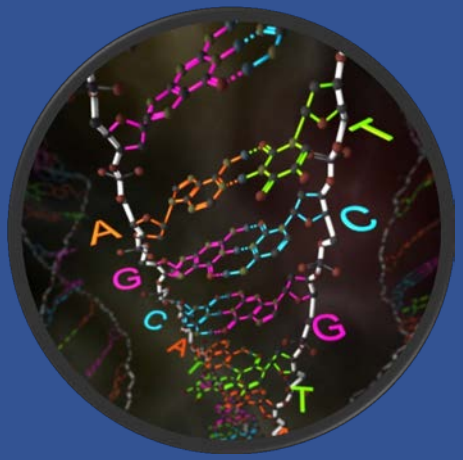




In contrast to some areas of physics, which require extremely expensive facilities, biology does not have an obvious need for 'big science.' Our country's spectacular success in this area has depended in large part on the wide support of independent, investigator-initiated, peer-reviewed research.

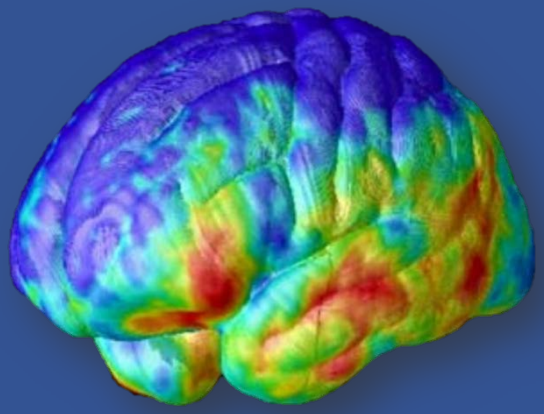
Now ***or*** **Then**

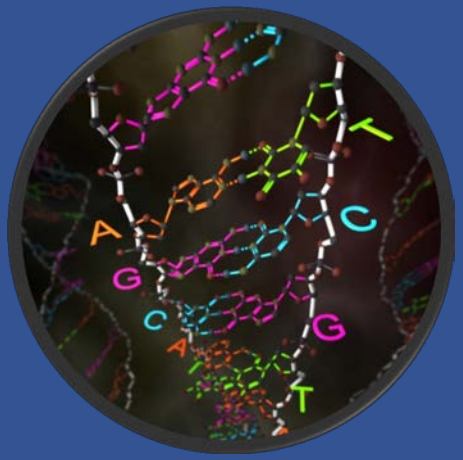




Creative science is bottom-up, not top-down. Are we talking about central planning inside the Beltway?

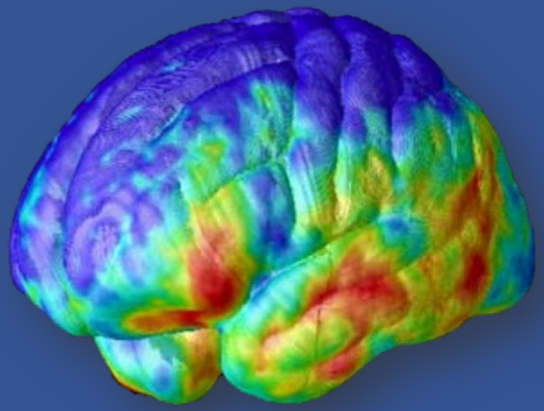
Now ***or*** **Then**

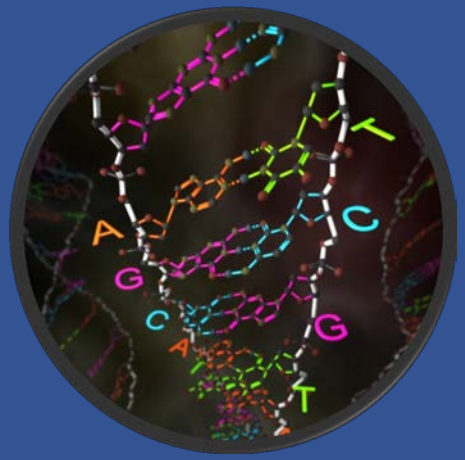




It’s going to do absolutely no good to develop tools for a new generation of [scientists] if we in the process seriously damage that same generation of scientists.

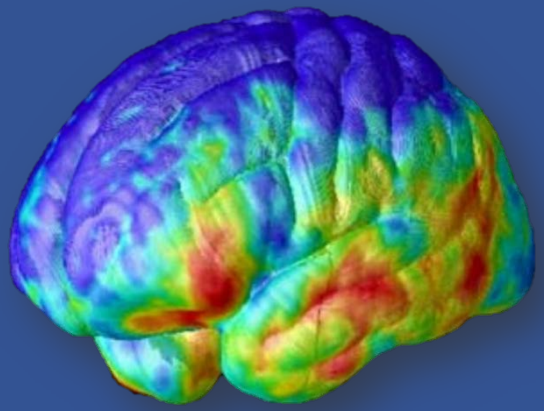
Now ***or*** **Then**

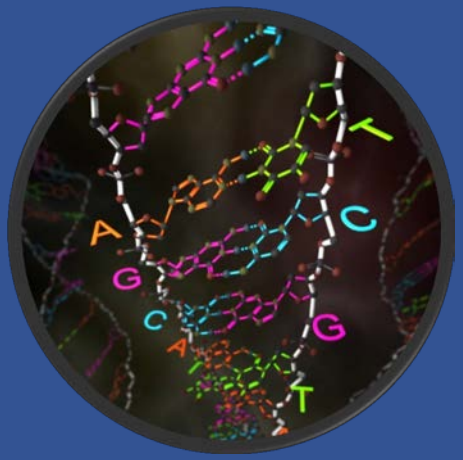




Arguments are made that the ? project will give birth to a new generation of technologies. What good will that do in the absence of individuals trained and capable of applying these technologies?

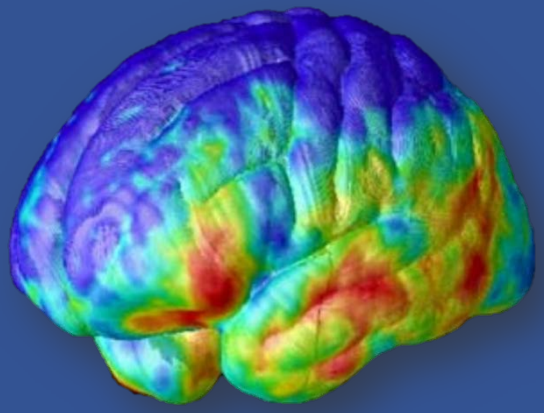
Now ***or*** **Then**

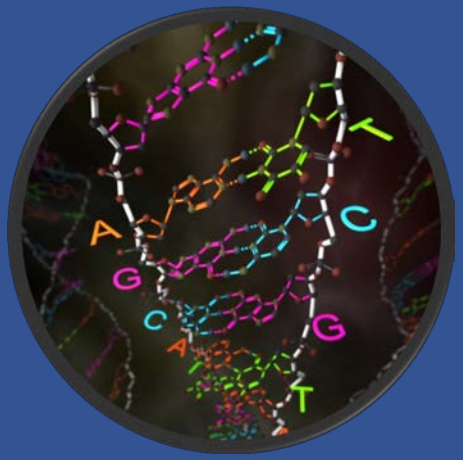




The amount of money we ask to accomplish the task, $200 million a year, is commensurate with the project's role in the fight against many serious health problems.

Now *or* **Then**



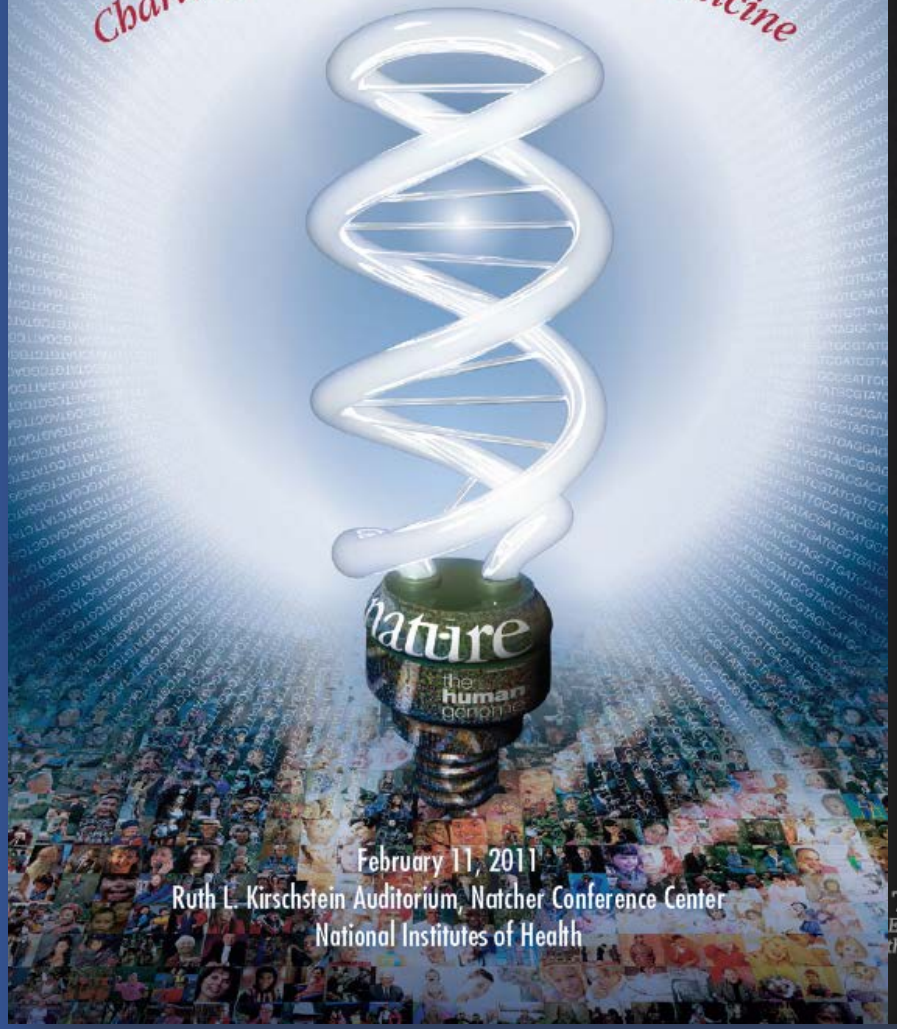


Everybody I talk to thinks this is an incredibly bad idea.

An NHGRI Symposium

A Decade with the Human Genome Sequence

Charting a Course for Genomic Medicine



February 11, 2011
Ruth L. Kirschstein Auditorium, Natcher Conference Center
National Institutes of Health

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

THE FUTURE IS BRIGHT

Reflections on the first ten years of the human genomics age

GENOMICS
THE END OF THE BEGINNING
Eric Lander on the impact of the human genome sequence
PAGE 187

METHODS
MORE BASES PER DOLLAR
Elaine Mardis on the march of sequencing technology
PAGE 198

HEALTH
FROM LAB TO CLINIC
A road map to genomic medicine
PAGE 204

NATURE.COM/NATURE
10 February 2011
Vol. 470, No. 7333

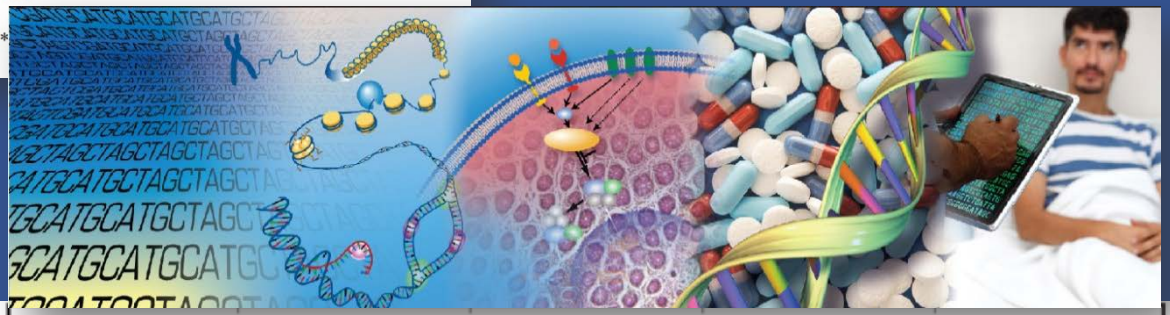
nature

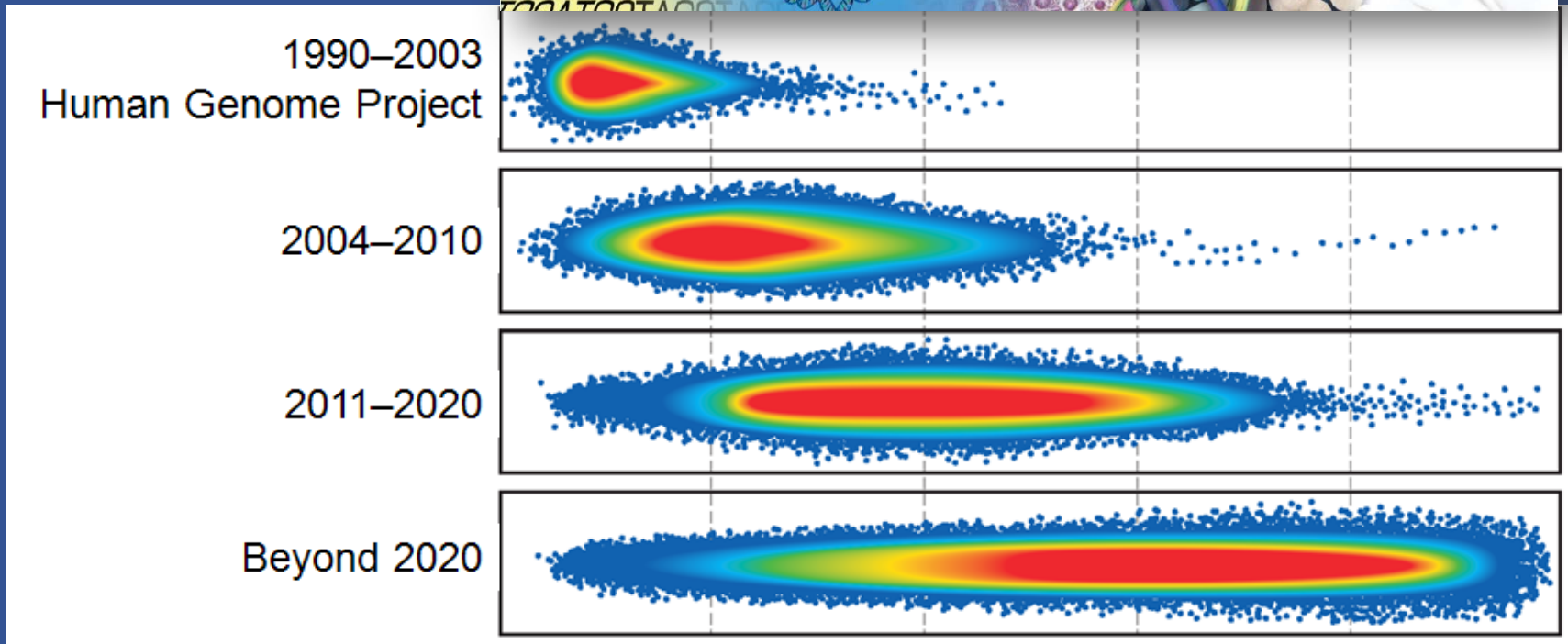
PERSPECTIVE

VOL 470 | 10 FEBRUARY 2011

Charting a course for genomic medicine from base pairs to bedside

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





Charting a Course for Genomic Medicine: Goals for the Next Ten Years

- Make genomics-based diagnostics routine
- Define genetic components of disease
- Complete comprehensive characterization of cancer genomes
- Devise practical systems for clinical genomic informatics
- Uncover role of human microbiome in health, disease

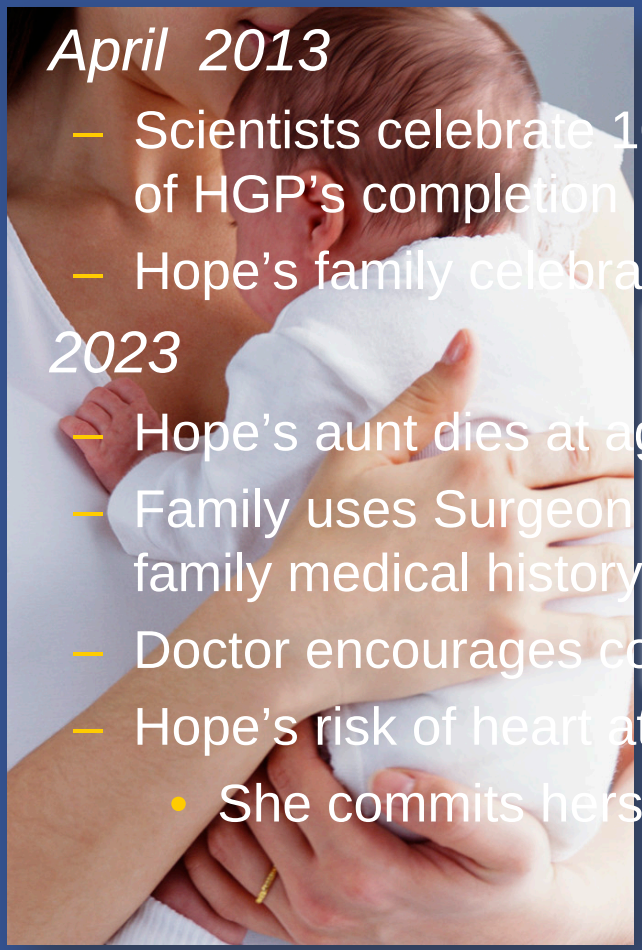
Achieving Our Goals by 2023

- Understanding the biology of genomes, with
  - New, comprehensive catalogues of data
  - New tools for genomics research
  - Rapid, internationally harmonized data release
- Bioinformatics and computational biology
  - Data analysis, integration
  - Visualization
- Education and training
  - Outreach to public, healthcare providers
  - Preparing next generation of researchers
- Genomics and society, addressing
  - Psychosocial, ethical issues
  - Legal, policy issues

Personalized Medicine: A future dream

Story of Hope

- *April 14, 2003:* baby Hope is born
  - Family hears news of HGP completion
- *April 2013*
  - Scientists celebrate 10th anniversary of HGP's completion
  - Hope's family celebrates her 10th
- *2023*
  - Hope's aunt dies at ag… failu…
  - Family uses Surgeon… family medical history
  - Doctor encourages co… e ar…
  - Hope's risk of heart attack 3x > than nor…
    - She commits herself to a healthy lifes…



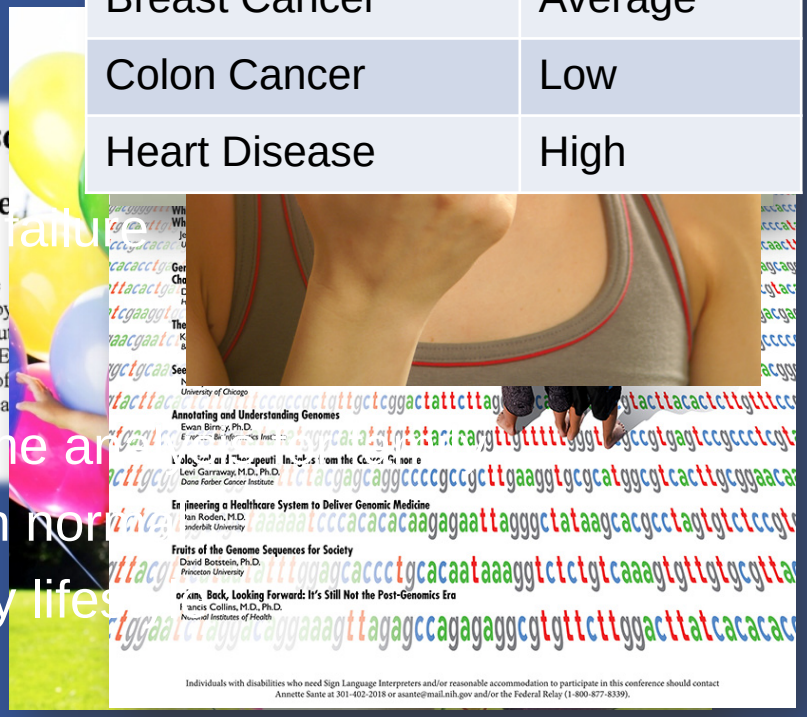
Trait	Risk
Alzheimer's Disease	Average
Breast Cancer	Average
Colon Cancer	Low
Heart Disease	High


International Cons…

All Goals Achieved; Ne…

BETHESDA, Md., April 14, 2003

The International Human Genome … Consortium … the United States by … al Human Genome Research Institu… and the Department of Energy (DOE … announced the successful completion of … Genome Project more than two yea… schedule.



Story of Hope

- *2053*
  - A healthy Hope celebrates her 50th
    - Wears “smart shirt” regularly
  - Hope & family also celebrate HGP’s 50th
- *2071*
  - Hope feels tightness in arm, assumes pulled muscle
  - Smart shirt calls emergency responders
  - Personalized drug therapy saves Hope

Story of Hope

- *2103*
  - Hope celebrates her 100th with a night of dancing

Personalized Medicine: Could the dream become a nightmare?

Hope’s Story Gone Wrong

- *2023:* Hope’s aunt, 53, dies of a heart attack
  - No online tool for family medical history available
  - Hope’s doctor dismisses genome analysis
  - Unaware of elevated heart disease risk, Hope makes unhealthy life choices
- *2038:* Hope diagnosed with high blood pressure
  - Pharmacogenomics *would have* determined best drug for her
  - Instead receives drug that causes hypersensitivity reaction; stops treatment

Hope's Story Gone Wrong – continued

- *2038–2052:* Hope eats an unhealthy diet, gains weight
- *2053:* Hope, gardening, feels tightness in her arm
  - No smart shirt; doctor dismisses it as pulled muscle
  - Taken to ER 3 hours later in cardiogenic shock
    - Given "standard" therapy – unable to metabolize prodrug
    - *Hope dies. She was just 50 years old.*

The essential goal of genomic medicine

Keep Hope Alive!




NIH...

Turning Discovery Into Health®