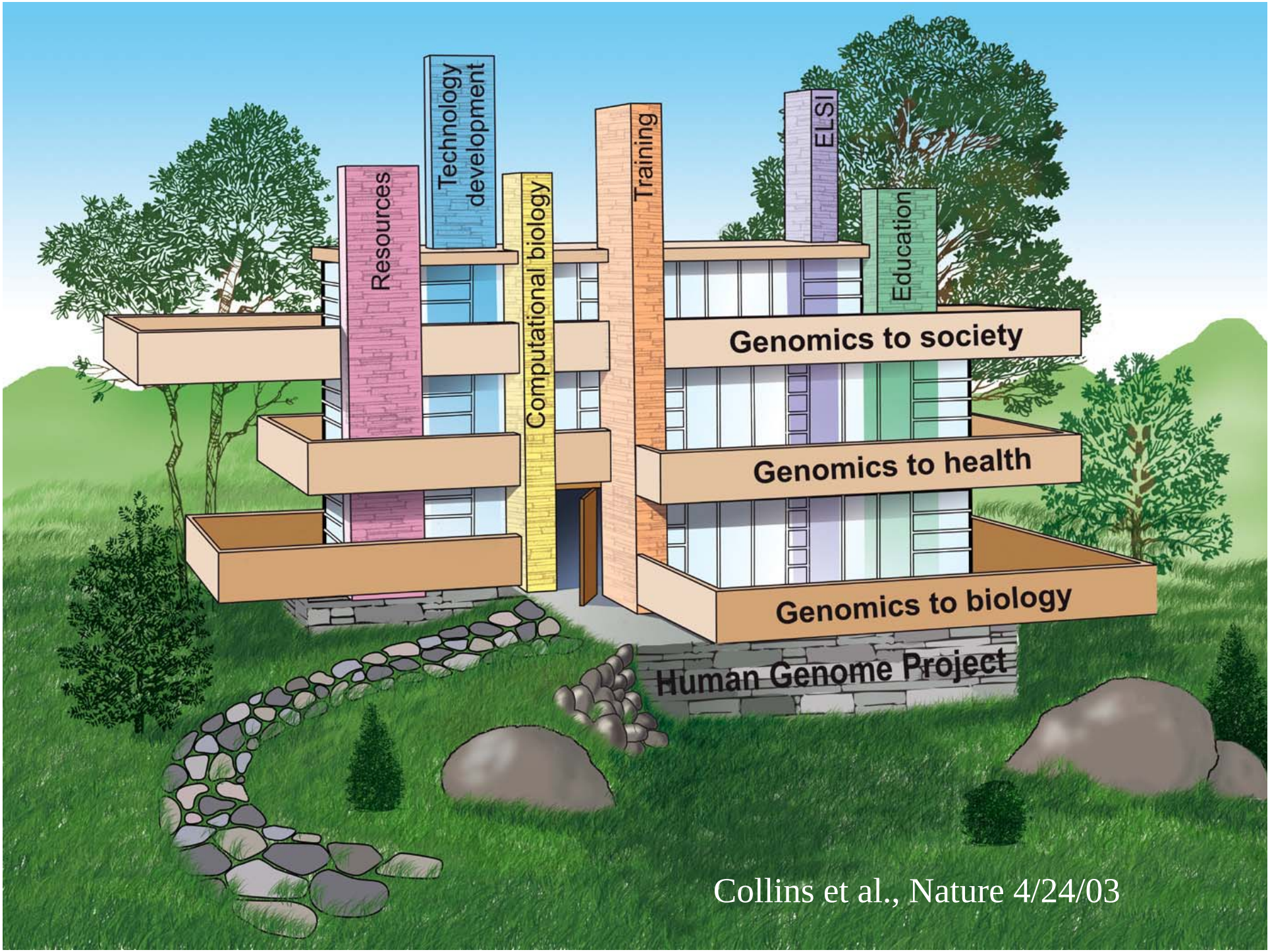
Population Genomics at NHGRI: History and Need

Frontiers in Population Genomics Workshop

Francis S. Collins, M.D., Ph.D.

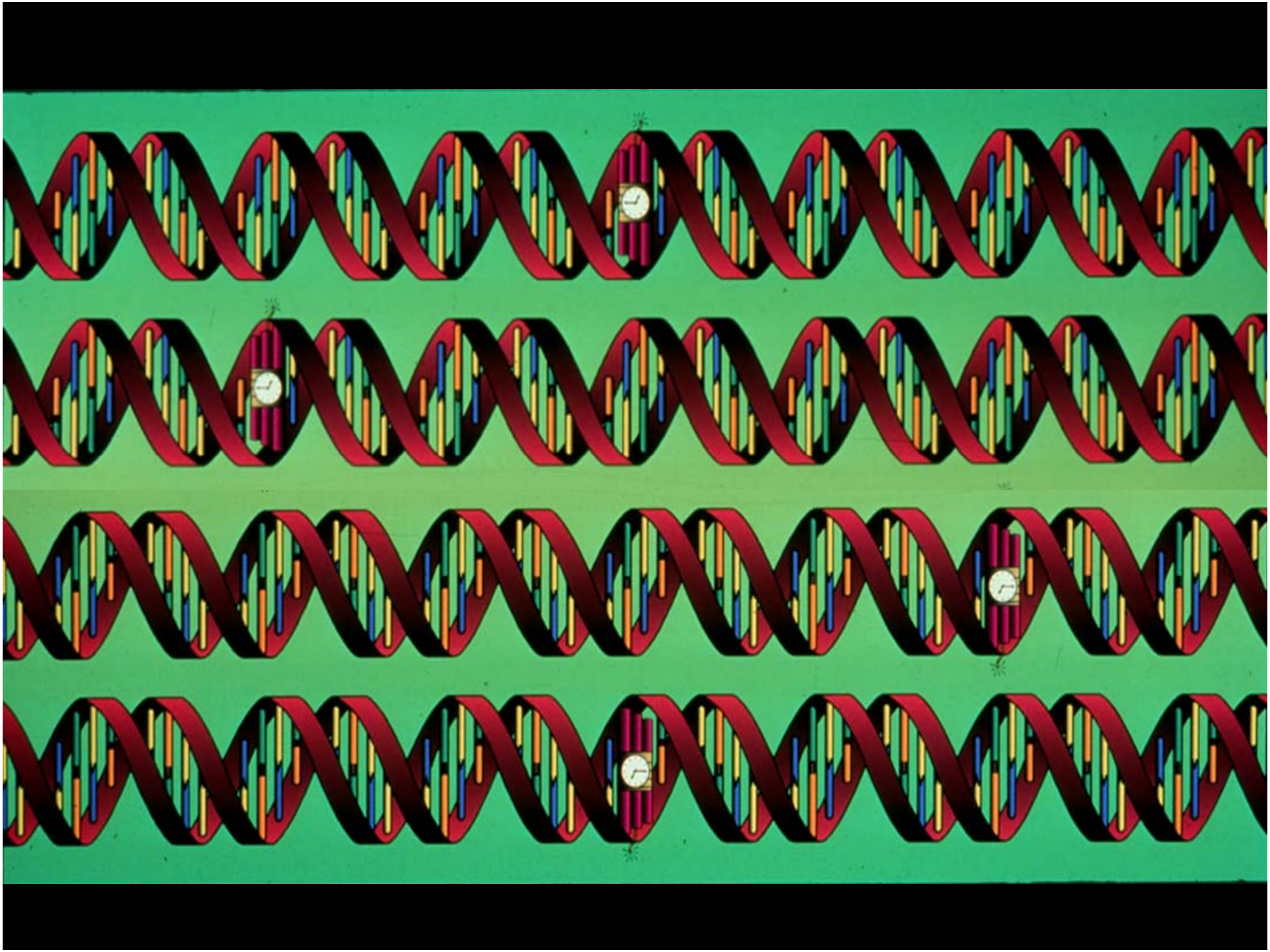
December 18, 2007


SALLY FORTH
HOW MUCH LONGER, TED?
FINDING THE RIGHT CHRISTMAS TREE TAKES TIME, RESEARCH AND RIGOROUS TESTING, SAL.
IT'S A TREE, TED, NOT THE GENOME PROJECT.
OKAY, THEN YOU PICK IT.
©2007 by King Features Syndicate, Inc. World rights reserved.


Resources
Technology development
Computational biology
Training
ELSI
Education
Genomics to society
Genomics to health
Genomics to biology
Human Genome Project
Collins et al., Nature 4/24/03

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics



"Genome Wide Association" Approach to Common Disease: The View from 2003

- Identify all 10 million common SNPs
- Collect 1000 cases and 1000 controls
- Genotype all DNAs for all SNPs
- That adds up to 20 billion genotypes
- At 50 cents a genotype, that's $10 billion for each disease – completely out of the question

27 October 2005 | www.nature.com/nature

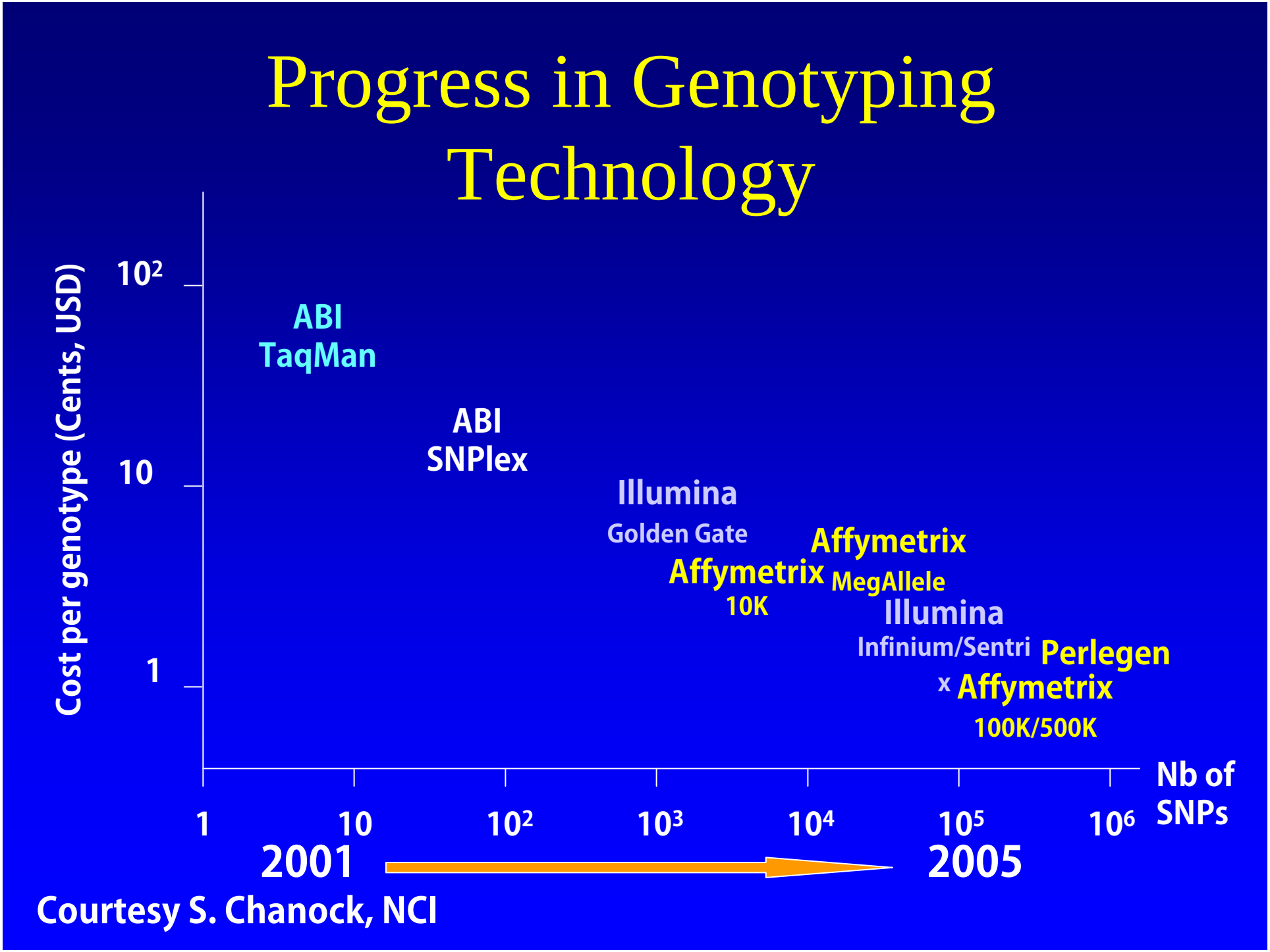
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature

THE HAPMAP PROJECT

Chapter and verse on human genetic variation

Progress in Genotyping Technology



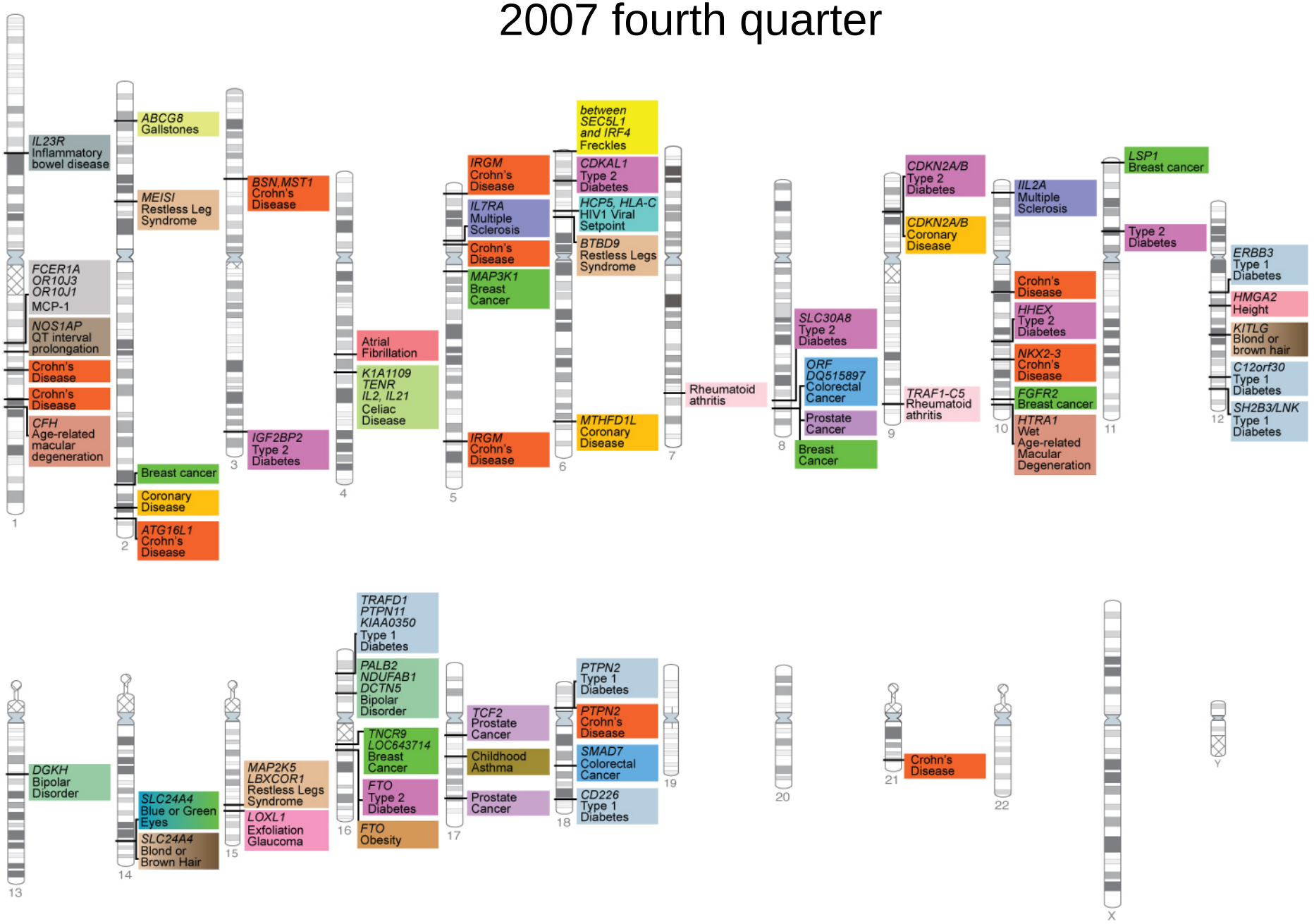
Courtesy S. Chanock, NCI

Genome Wide Association Approach to Common Disease: The View from 2007

- Identify an optimum set of 300,000 tag SNPs
- Collect 1000 cases and 1000 controls
- Genotype all DNAs for all SNPs
- That adds up to 600 million genotypes
- Genotyping just dropped to $0.0012, so that's $800,000 for each disease

GENETIC ASSOCIATION INFORMATION NETWORK

2007 fourth quarter



2007: Genome-wide association works!

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing

HUMAN GENOMIC SEQUENCE GENERATED BY LARGE SCALE CENTRES

RELEASE

- Automatic release of sequence as available >1 kb (preferably daily)
- Immediate submission of finished annotated sequence

[redacted] and in public domain

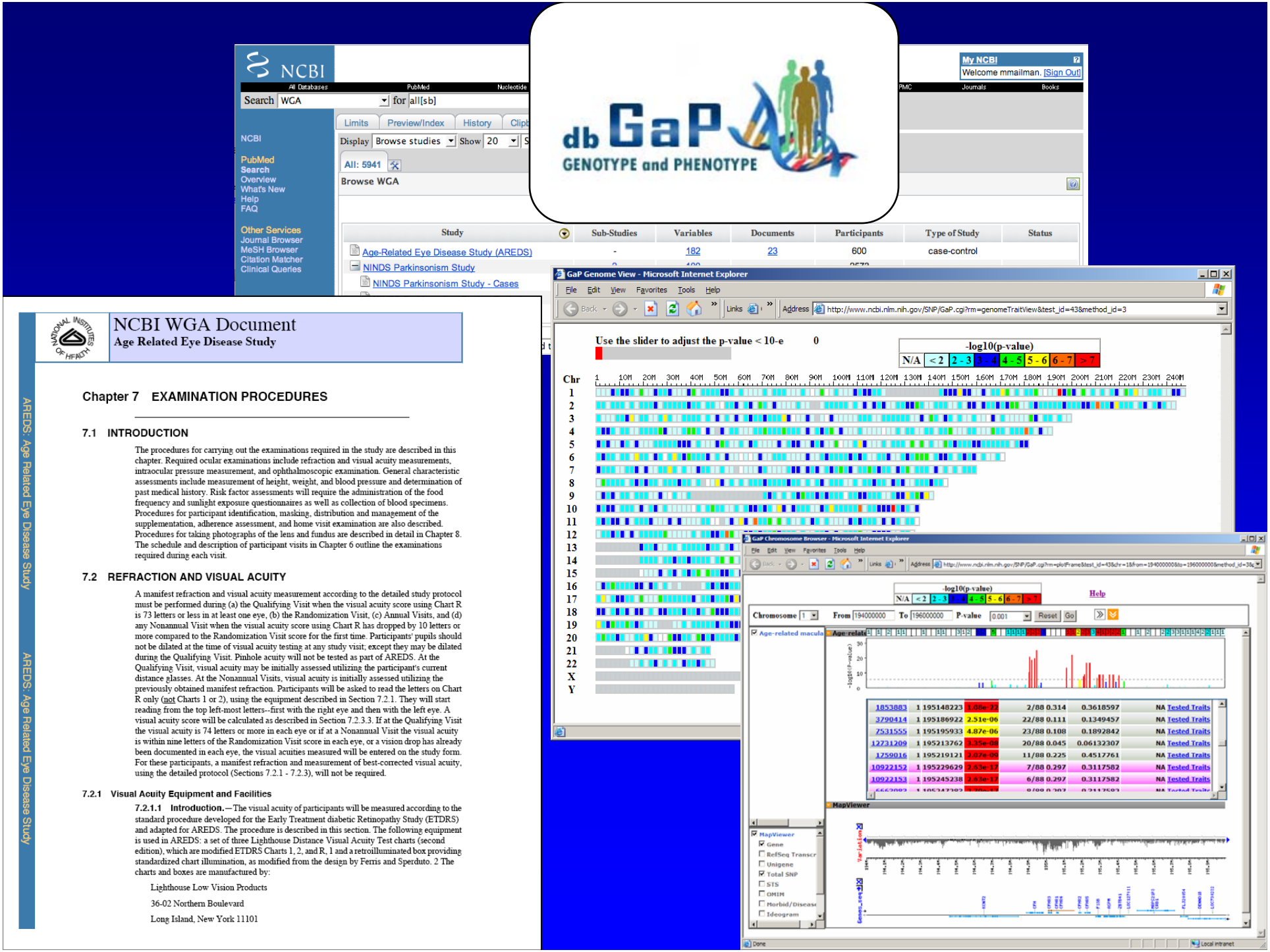
- Aim to have all sequence freely available for both research and development, in order to maximise its benefits to society.

POLICY

- The funding agencies are urged to foster these policies

NIH Guiding Principle

The greatest public benefit will be realized if data from GWAS are made available, under terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of investigators.



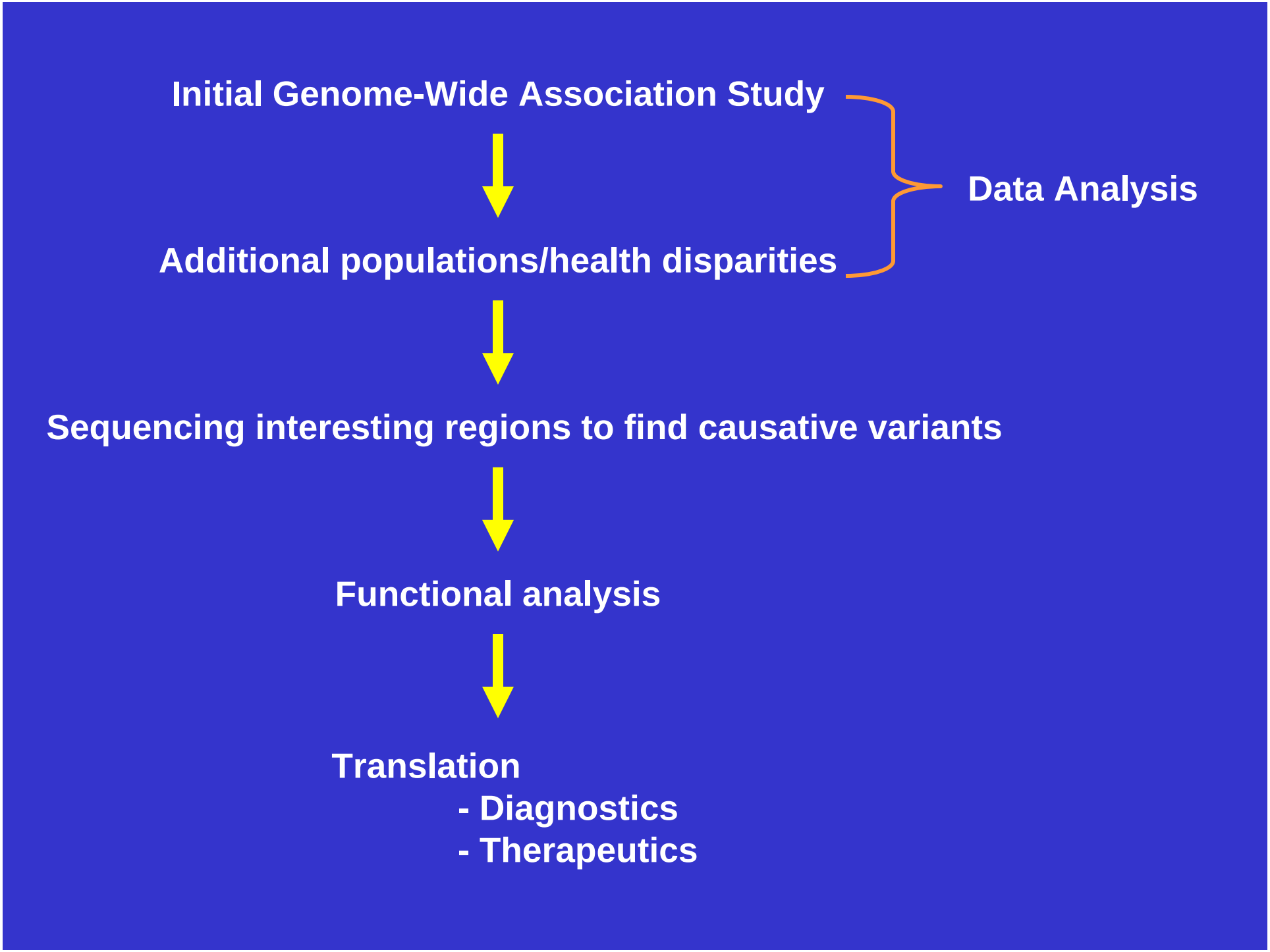
The procedures for carrying out the examinations required in the study are described in this chapter. Required ocular examinations include refraction and visual acuity measurements, intraocular pressure measurement, and ophthalmoscopic examination. General characteristic assessments include measurement of height, weight, and blood pressure and determination of past medical history. Risk factor assessments will require the administration of the food frequency and sunlight exposure questionnaires as well as collection of blood specimens. Procedures for participant identification, masking, distribution and management of the supplementation, adherence assessment, and home visit examination are also described. Procedures for taking photographs of the lens and fundus are described in detail in Chapter 8. The schedule and description of participant visits in Chapter 6 outline the examinations required during each visit.

A manifest refraction and visual acuity measurement according to the detailed study protocol must be performed during (a) the Qualifying Visit when the visual acuity score using Chart R is 73 letters or less in at least one eye, (b) the Randomization Visit, (c) Annual Visits, and (d) any Nonannual Visit when the visual acuity score using Chart R has dropped by 10 letters or more compared to the Randomization Visit score for the first time. Participants' pupils should not be dilated at the time of visual acuity testing at any study visit; except they may be dilated during the Qualifying Visit. Pinhole acuity will not be tested as part of AREDS. At the Qualifying Visit, visual acuity may be initially assessed utilizing the participant's current distance glasses. At the Nonannual Visits, visual acuity is initially assessed utilizing the previously obtained manifest refraction. Participants will be asked to read the letters on Chart R only (not Charts 1 or 2), using the equipment described in Section 7.2.1. They will start reading from the top left-most letters--first with the right eye and then with the left eye. A visual acuity score will be calculated as described in Section 7.2.3.3. If at the Qualifying Visit the visual acuity is 74 letters or more in each eye or if at a Nonannual Visit the visual acuity is within nine letters of the Randomization Visit score in each eye, or a vision drop has already been documented in each eye, the visual acuities measured will be entered on the study form. For these participants, a manifest refraction and measurement of best-corrected visual acuity, using the detailed protocol (Sections 7.2.1 - 7.2.3), will not be required.

7.2.1.1 Introduction.—The visual acuity of participants will be measured according to the standard procedure developed for the Early Treatment diabetic Retinopathy Study (ETDRS) and adapted for AREDS. The procedure is described in this section. The following equipment is used in AREDS: a set of three Lighthouse Distance Visual Acuity Test charts (second edition), which are modified ETDRS Charts 1, 2, and R, 1 and a retroilluminated box providing standardized chart illumination, as modified from the design by Ferris and Sperduto. 2 The charts and boxes are manufactured by:

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics


Initial Genome-Wide Association Study
Additional populations/health disparities
Data Analysis
Sequencing interesting regions to find causative variants
Functional analysis
Translation
- Diagnostics
- Therapeutics

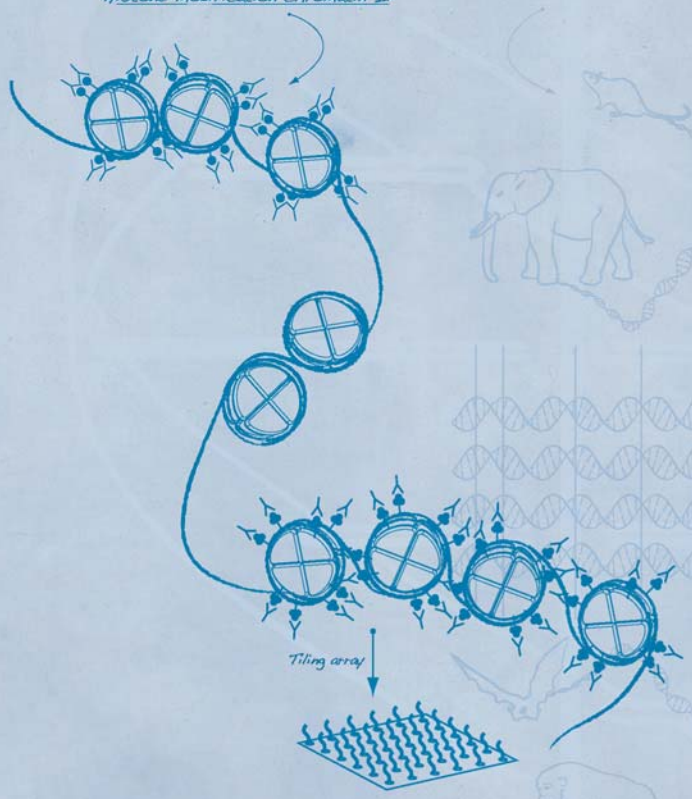
The Multiple Genomes Project (MultiGen)

- Plan is to sequence ~1000 genomes from 11 different populations at ~2 - 4x coverage, over 2 years, with new technology, and immediate public release of data
- Will go even deeper in exons
- Goal is to identify nearly all variants with frequency of the minor allele greater than 1% (down to 0.2% in exons)

14 June 2007 | www.nature.com/nature | $10

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature



MARS'S ANCIENT OCEAN
Polar wander solves an enigma

THE DEPTHS OF DISGUST
Understanding the ugliest emotion

MENTORING
How to be top

NATUREJOBS
Contract research

DECODING THE BLUEPRINT

The ENCODE pilot maps human genome function

$10.00US $12.99CAN
24
0 71486 03070 6

Significant Events in Population Genomics 2003 - 2007

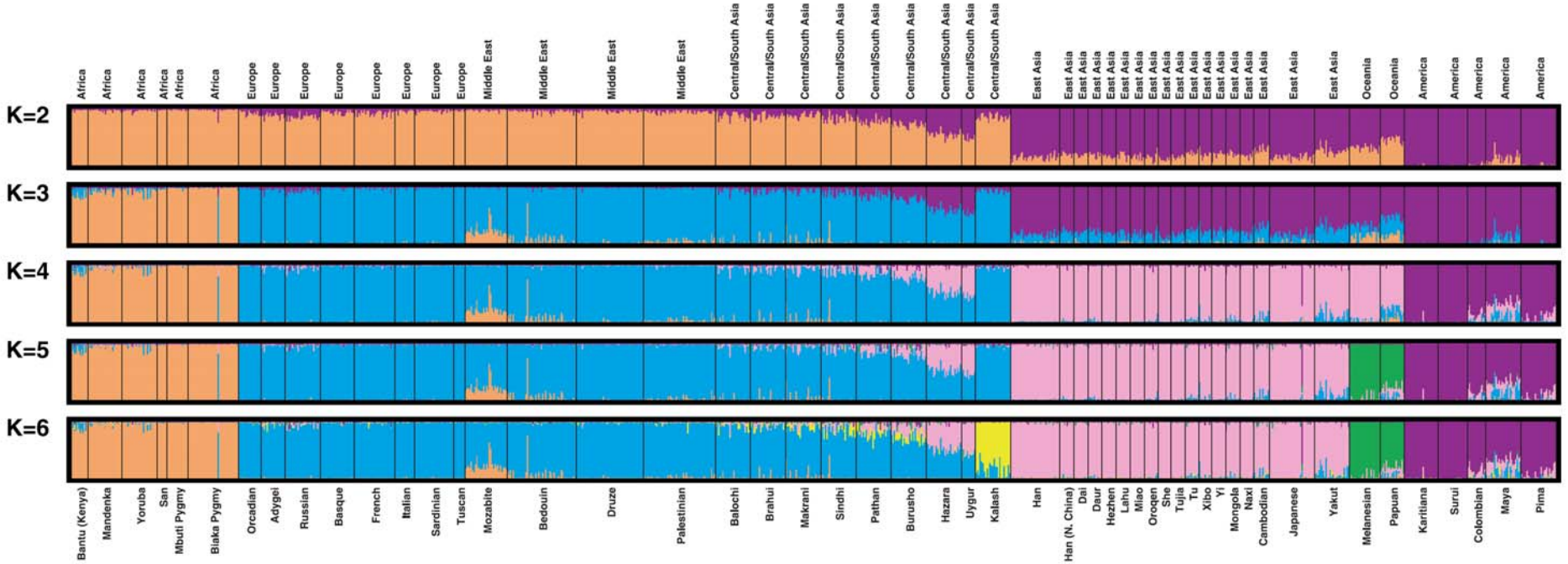
- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics
- **Early indications that genomics may be highly relevant to health disparities in common disease**

Genetic Structure of Human Populations

Noah A. Rosenberg,[1]* Jonathan K. Pritchard,[2] James L. Weber,[3] Howard M. Cann,[4] Kenneth K. Kidd,[5] Lev A. Zhivotovsky,[6] Marcus W. Feldman[7]

We studied human population structure using genotypes at 377 autosomal microsatellite loci in 1056 individuals from 52 populations. Within-population differences among individuals account for 93 to 95% of genetic variation; differences among major groups constitute only 3 to 5%. Nevertheless, without using prior information about the origins of individuals, we identified six main genetic clusters, five of which correspond to major geographic regions, and subclusters that often correspond to individual populations. General agreement of genetic and predefined populations suggests that self-reported ancestry can facilitate assessments of epidemiological risks but does not obviate the need to use genetic information in genetic association studies.

SCIENCE VOL 298 20 DECEMBER 2002



A common variant associated with prostate cancer in European and African populations

Laufey T Amundadottir[1,12], Patrick Sulem[1,12], Julius Gudmundsson[1,12], Agnar Helgason[1], Adam Baker[1],

Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24

Multiple regions within 8q24 independently affect risk for prostate cancer

Christopher A Haiman[1], Nick Patterson[2], Matthew L Freedman[2,3], Simon R Myers[2], Malcolm C Pike[1],

Genome-wide association study of prostate cancer identifies a second risk locus at 8q24

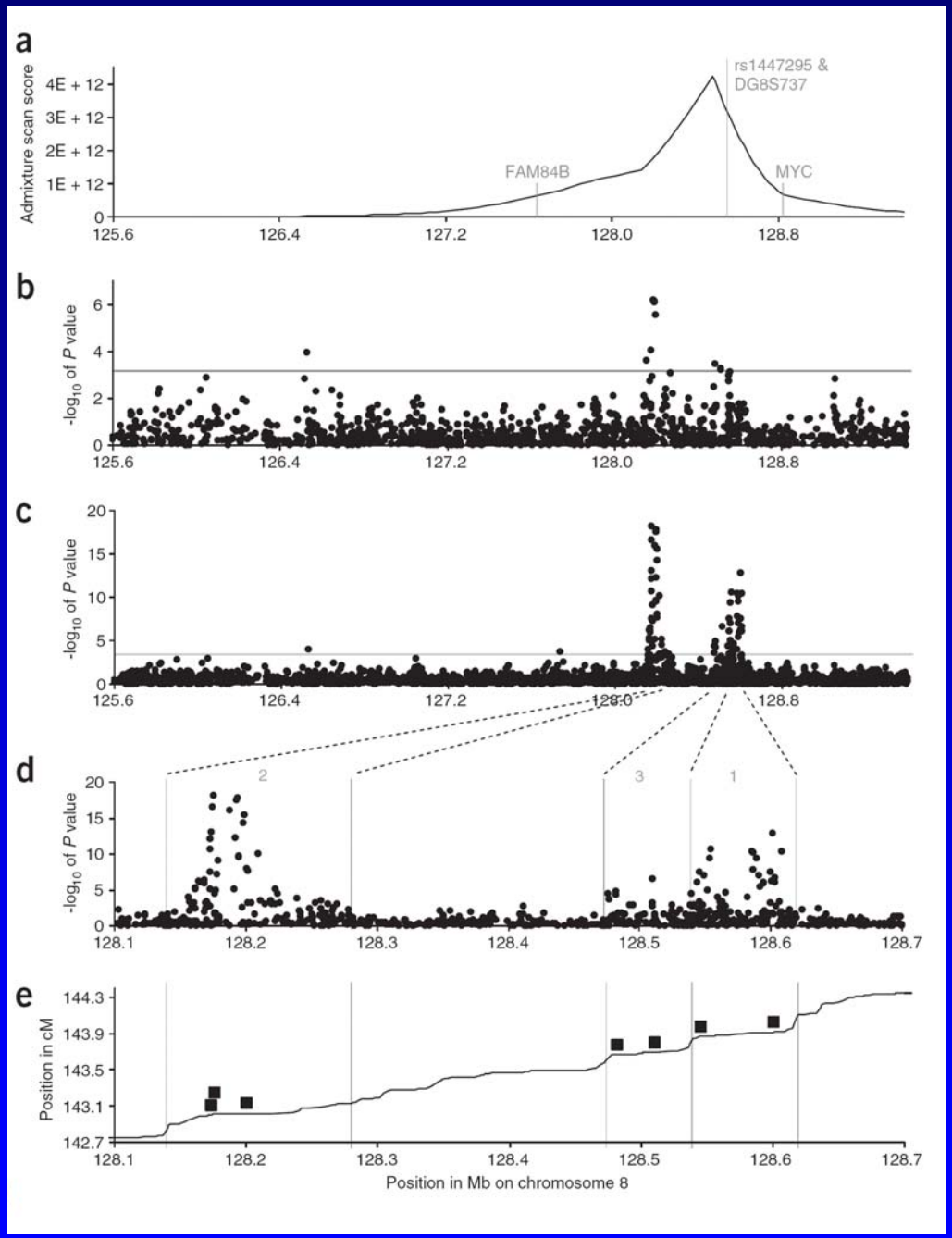
Meredith Yeager[1,2], Nick Orr[3], Richard B Hayes[2], Kevin B Jacobs[4], Peter Kraft[5], Sholom Wacholder[2], Mark J Minichiello[6], Paul Fearnhead[7], Kai Yu[2], Nilanjan Chatterjee[2], Zhaoming Wang[1,2], Robert Welch[1,2], Brian J Staats[1,2], Eugenia E Calle[8], Heather Spencer Feigelson[8], Michael J Thun[8], Carmen Rodriguez[8], Demetrius Albanes[2], Jarmo Virtamo[9], Stephanie Weinstein[2], Fredrick R Schumacher[5], Edward Giovannucci[10], Walter C Willett[10], Geraldine Cancel-Tassin[11], Olivier Cussenot[11], Antoine Valeri[11], Gerald L Andriole[12], Edward P Gelmann[13], Margaret Tucker[2], Daniela S Gerhard[14], Joseph F Fraumeni Jr[2], Robert Hoover[2], David J Hunter[2,5], Stephen J Chanock[2,3] & Gilles Thomas[2]

Prostate Cancer

Association of DG8S737 "Allele – 8" with Prostate Cancer

	Allele Frequency			
	Cases	Controls	OR	P-value
Iceland	0.131	0.078	1.77	$2x10^{-8}$
Sweden	0.101	0.079	1.38	$4x10^{-3}$
Chicago (European)	0.082	0.041	2.10	$3x10^{-3}$
Michigan (African-American)	0.234	0.161	1.60	$2x10^{-3}$

Amundadottir et al., Nature Genetics 38:652-8, 2006



Haiman *et al.*

Nature Genetics

***39*: 638-44 (2007)**

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics
- Early indications that genomics may be highly relevant to health disparities in common disease
- **Appreciation of need for better assessment of E and GxE**

Genes and Environment Initiative

EXPOSURE BIOLOGY PROGRAM

Develop technology and biomarkers

- Diet
- Physical Activity
- Environmental Exposures
- Psychosocial Stress and Addictive Substances

GXE

GENETICS PROGRAM

Identify genetic variants

- GWA Studies
- Data Analysis
- Replication
- Sequencing
- Database
- Function
- Translation

Gene Environment Association (GENEVA) Studies

Principal Investigator	Institution	Primary Outcome
Beaty, Terri	Johns Hopkins U	Oral clefts
Bierut, Laura	Washington U	Alcohol addiction
Boerwinkle, Eric	U of Texas	Coronary heart disease
Caporaso, Neil	National Cancer Inst	Lung cancer
Hu, Frank	Harvard U	Type 2 diabetes
Lowe, William	Northwestern U	Birth weight/ maternal glycemia
Marazita, Mary	U of Pittsburgh	Dental caries
Murray, Jeffrey	U of Iowa	Premature birth

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics
- Early indications that genomics may be highly relevant to health disparities in common disease
- Appreciation of need for better assessment of E and GxE
- **Discussion, but not much resolution, of the need for large scale population cohort studies in the U.S.**

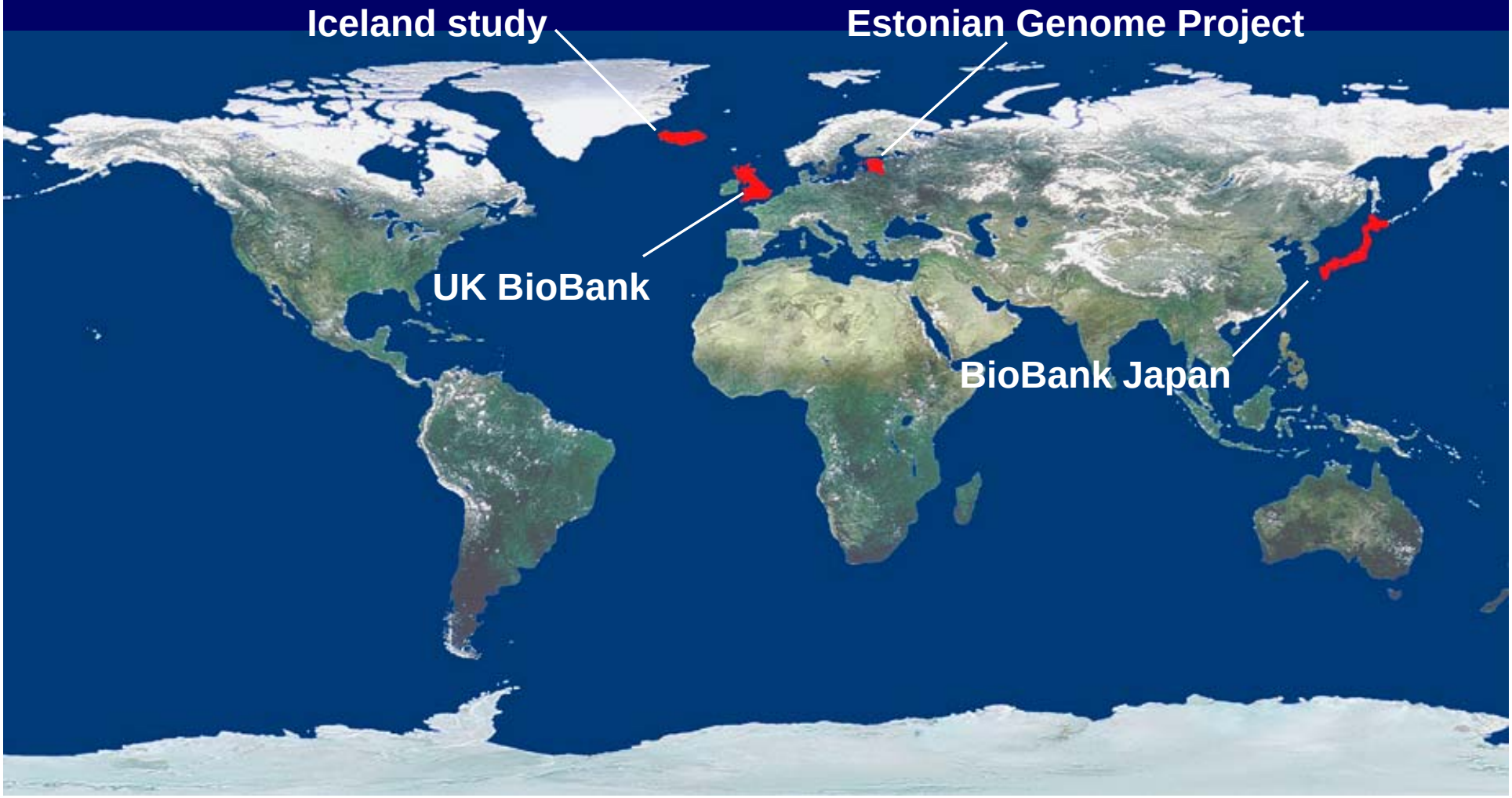
insight commentary

The case for a US prospective cohort study of genes and environment

Francis S. Collins

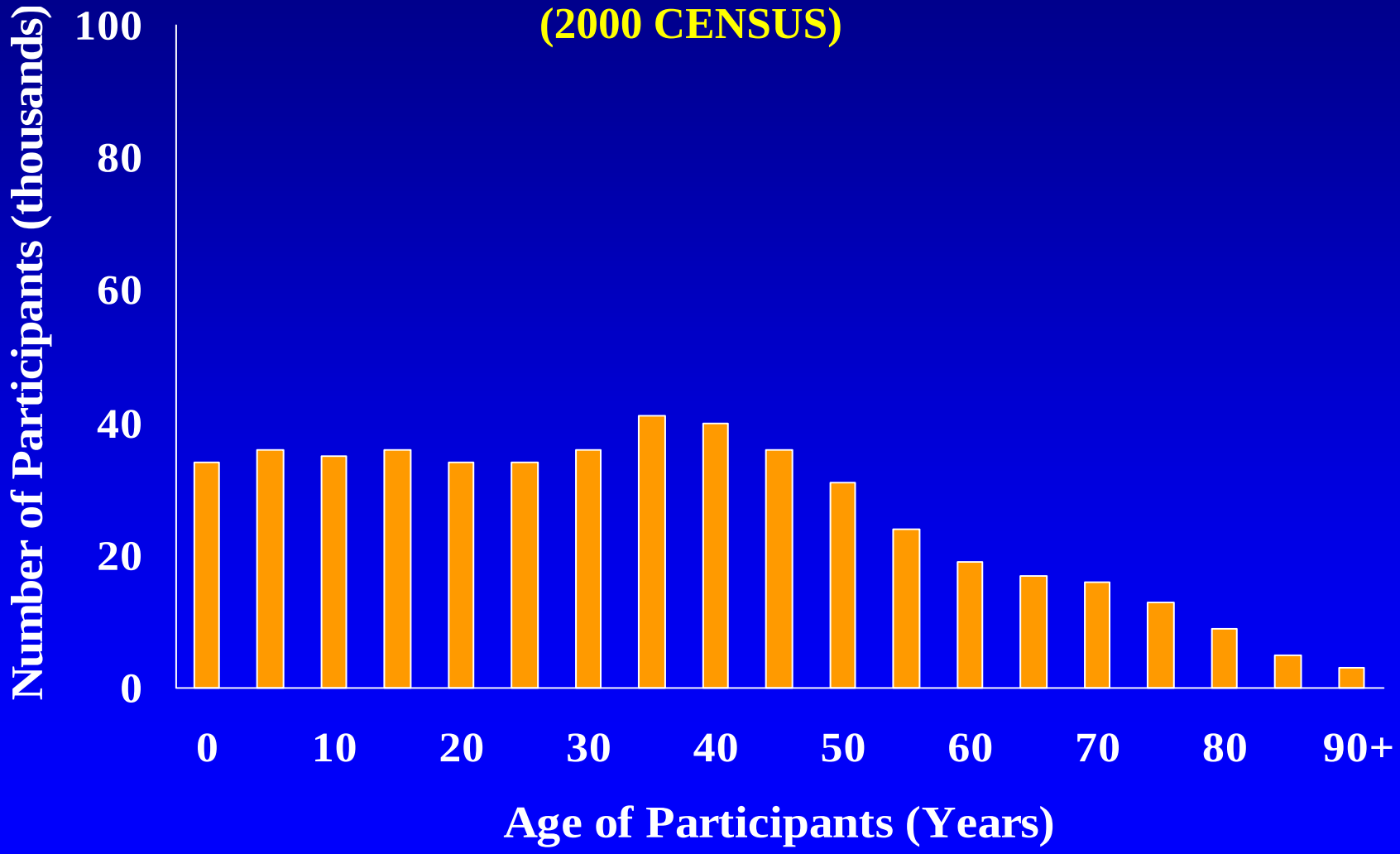
National Human Genome Research Institute, National Institutes of Health, Building 31, Room 4B09, MSC 2152, 31 Center Drive, Bethesda, Maryland 20892-2152, USA (e-mail: fc23a@nih.gov)

Other countries are planning large population studies of genes, environment, and health -- but those will not provide easy access to U.S. investigators, nor address U.S. health disparities, environmental exposures, or emerging health issues

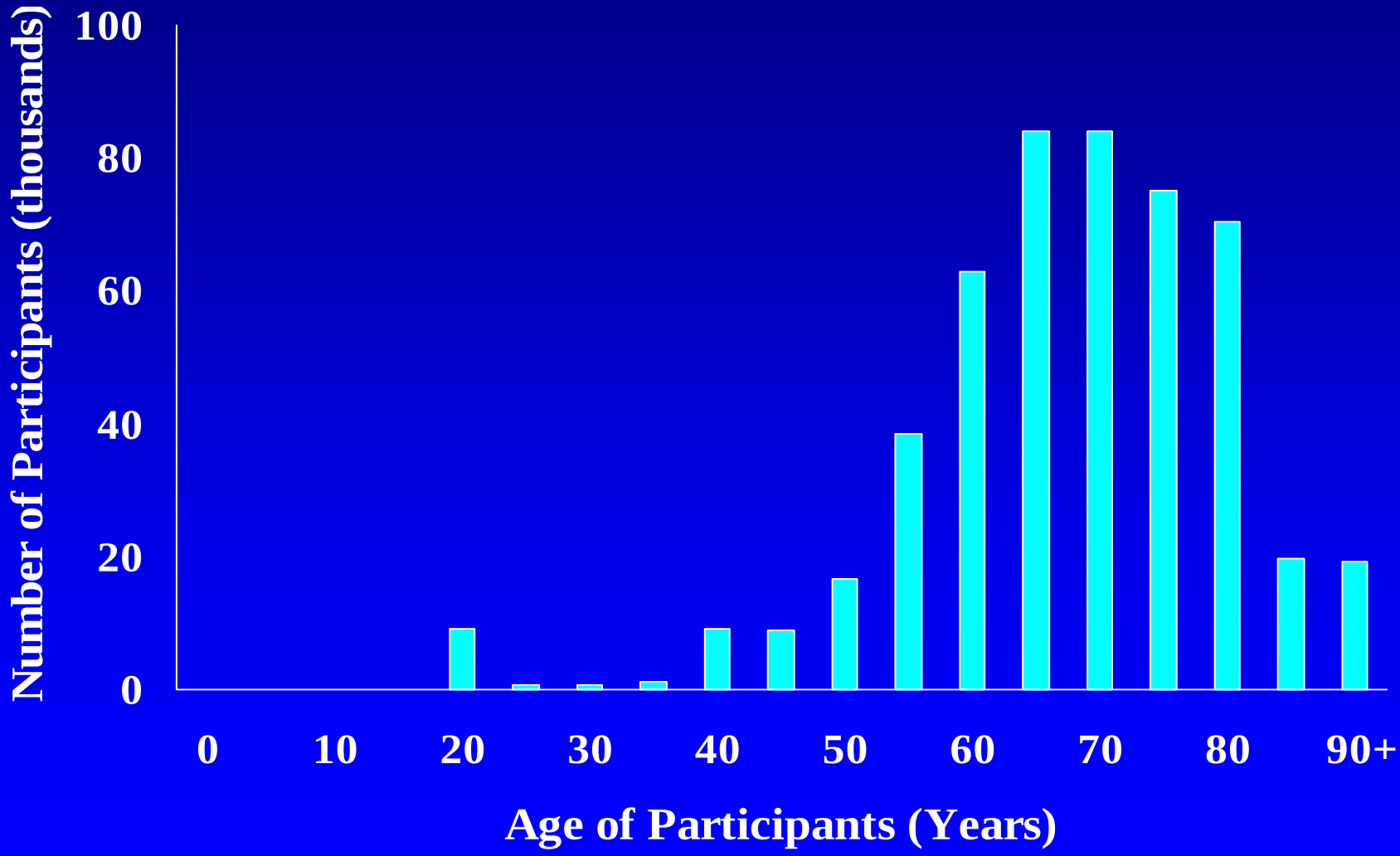


ESTIMATED AGE DISTRIBUTION OF REPRESENTATIVE US COHORT

(2000 CENSUS)

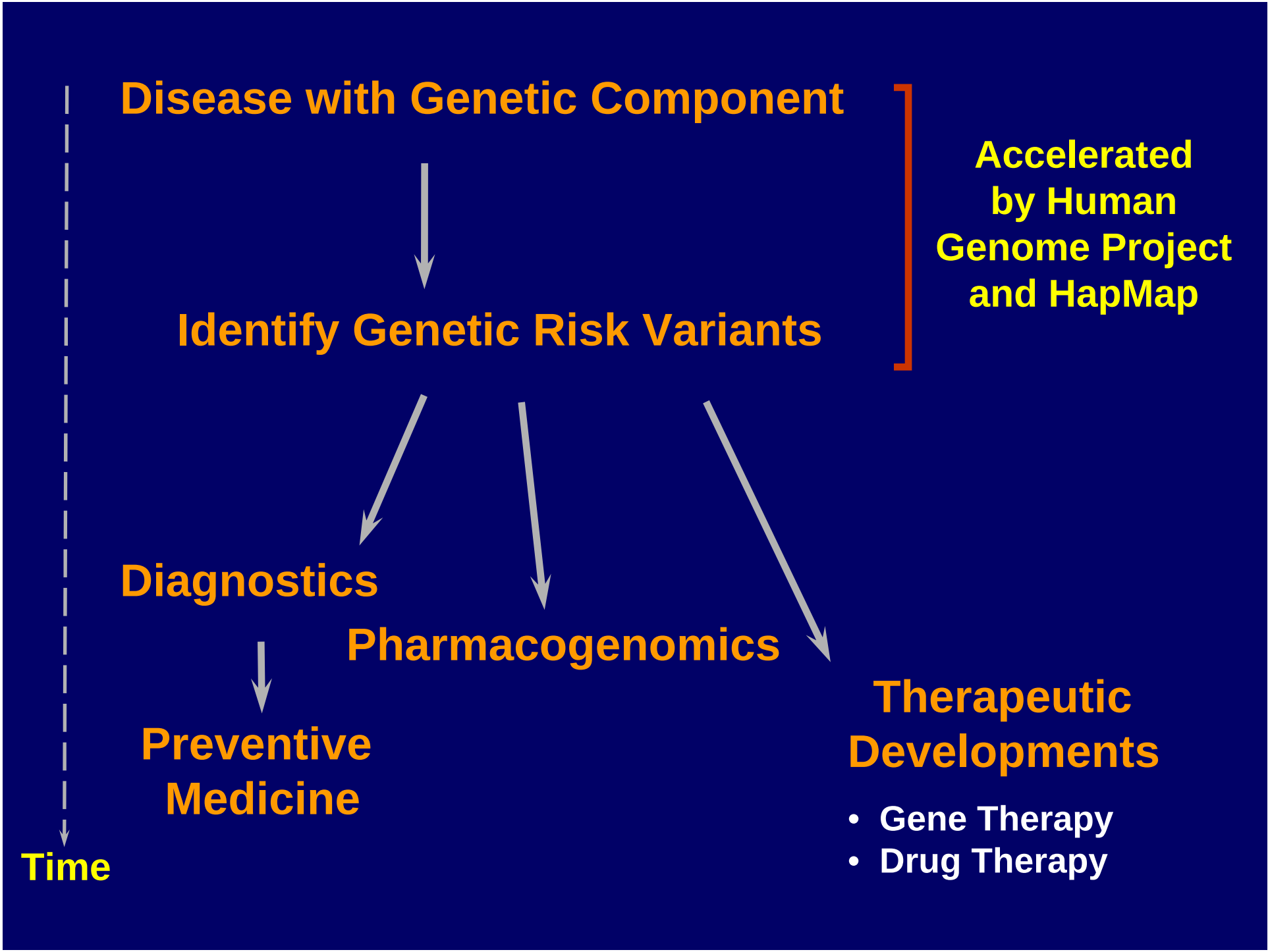


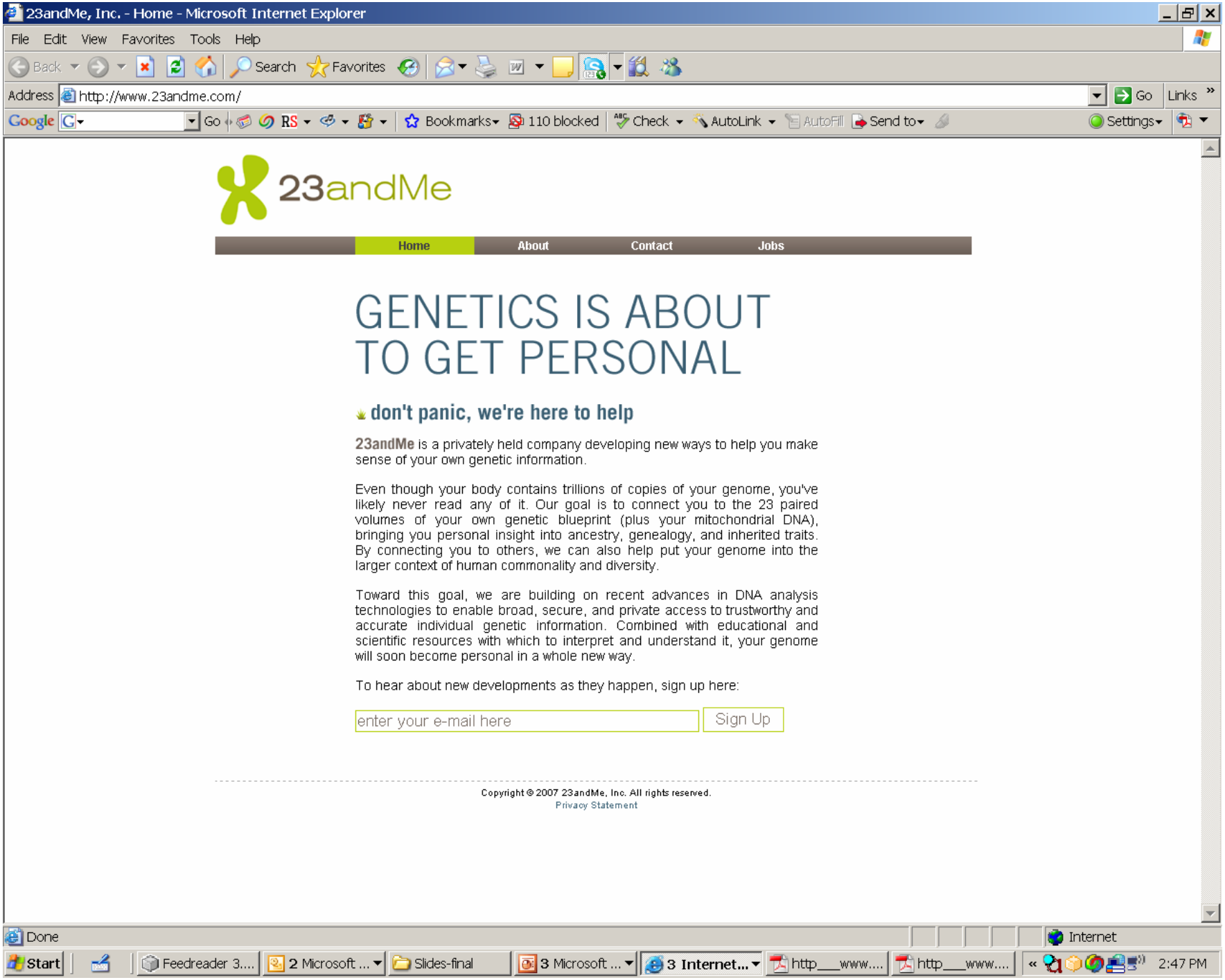
ESTIMATED AGE DISTRIBUTION OF EXISTING NIH-FUNDED COHORTS

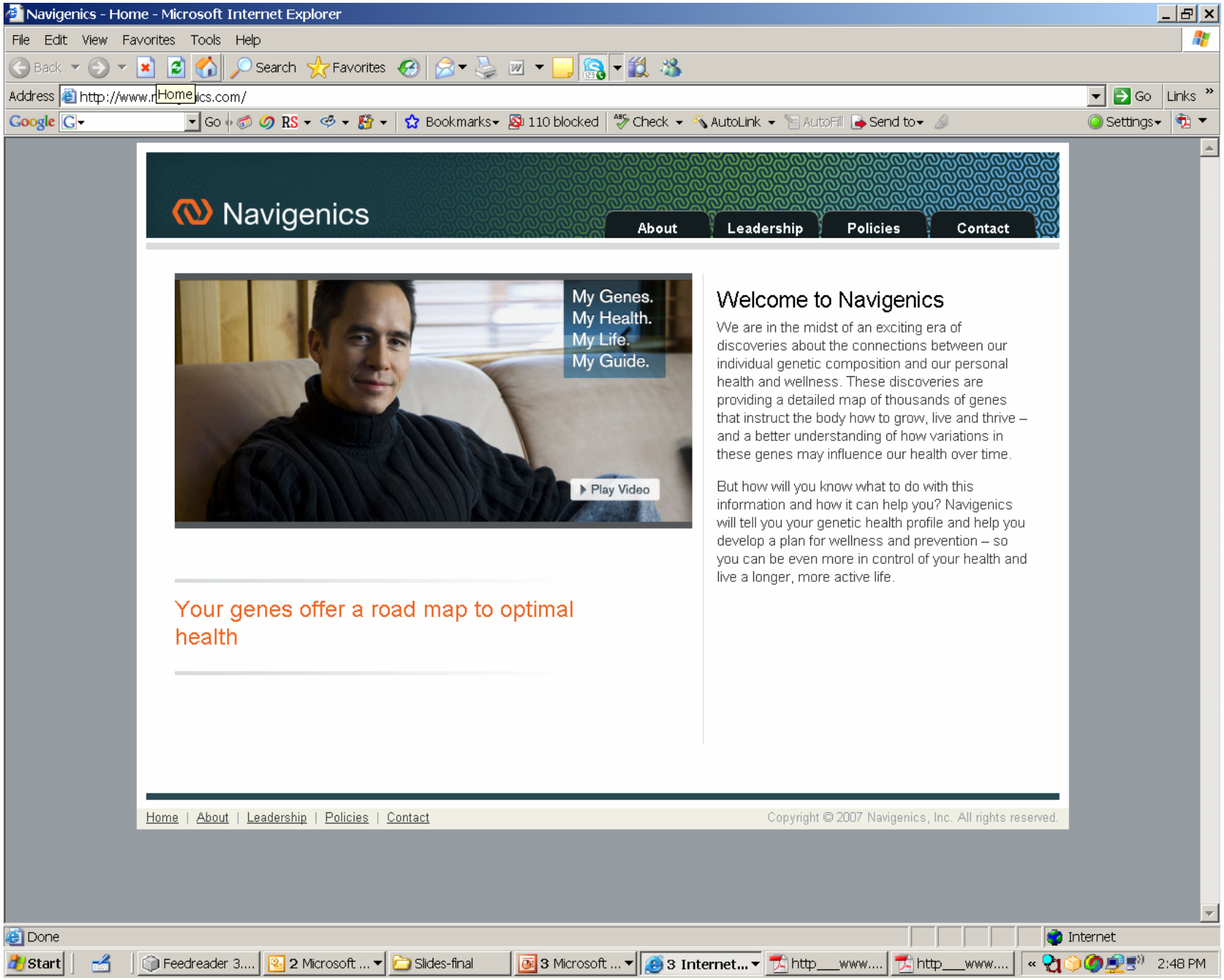


Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics
- Early indications that genomics may be highly relevant to health disparities in common disease
- Appreciation of need for better assessment of E and GxE
- Discussion, but not much resolution, of the need for large scale population cohort studies in the U.S.
- **Introduction of genetic risk factor assessment into mainstream medicine – too soon?**


Disease with Genetic Component
Identify Genetic Risk Variants
Accelerated by Human Genome Project and HapMap
Diagnostics
Pharmacogenomics
Therapeutic Developments
Preventive Medicine
• Gene Therapy
• Drug Therapy
Time


23andMe, Inc. - Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.23andme.com/
Go Links
Google Go Bookmarks 110 blocked Check AutoLink AutoFill Send to Settings
23andMe
Home About Contact Jobs
GENETICS IS ABOUT TO GET PERSONAL
don't panic, we're here to help
23andMe is a privately held company developing new ways to help you make sense of your own genetic information.
Even though your body contains trillions of copies of your genome, you've likely never read any of it. Our goal is to connect you to the 23 paired volumes of your own genetic blueprint (plus your mitochondrial DNA), bringing you personal insight into ancestry, genealogy, and inherited traits. By connecting you to others, we can also help put your genome into the larger context of human commonality and diversity.
Toward this goal, we are building on recent advances in DNA analysis technologies to enable broad, secure, and private access to trustworthy and accurate individual genetic information. Combined with educational and scientific resources with which to interpret and understand it, your genome will soon become personal in a whole new way.
To hear about new developments as they happen, sign up here:
enter your e-mail here
Sign Up
Copyright © 2007 23andMe, Inc. All rights reserved.
Privacy Statement
Done
Internet
Start
Feedreader 3....
2 Microsoft ...
Slides-final
3 Microsoft ...
3 Internet...
http___www....
http___www....
2:47 PM


Navigenics - Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.navigenics.com/
Home
Go Links
Google Go Bookmarks 110 blocked Check AutoLink AutoFill Send to Settings
Navigenics
About
Leadership
Policies
Contact
My Genes.
My Health.
My Life.
My Guide.
Play Video
Welcome to Navigenics
We are in the midst of an exciting era of discoveries about the connections between our individual genetic composition and our personal health and wellness. These discoveries are providing a detailed map of thousands of genes that instruct the body how to grow, live and thrive – and a better understanding of how variations in these genes may influence our health over time.
But how will you know what to do with this information and how it can help you? Navigenics will tell you your genetic health profile and help you develop a plan for wellness and prevention – so you can be even more in control of your health and live a longer, more active life.
Your genes offer a road map to optimal health
Home | About | Leadership | Policies | Contact
Copyright © 2007 Navigenics, Inc. All rights reserved.
Done
Internet
Start
Feedreader 3....
2 Microsoft ...
Slides-final
3 Microsoft ...
3 Internet...
http___www....
http___www....
2:48 PM


WHAT WENT WRONG IN IRAQ (HINT: BLAME THE GEEKS)
WIRED
your LIFE: decoded
A new $1,000 DNA test can tell you how you'll live—and die. Welcome to the Age of the Genome.
BY THOMAS GOETZ
WISH LIST
GADGETS AND GEAR FOR THE HOLIDAYS
HOW TO
MEMORIZE A 20-DIGIT NUMBER!
KNOW THYSELF | dec 2007

Study to Probe How Healthy Younger Adults Make Use of Genetic Tests

Multiplex Initiative Offers Genetic Testing to Participants in Metropolitan Detroit

BETHESDA, Md., Fri., May 4, 2007 — The National Human Genome Research Institute (NHGRI) and the National Cancer Institute (NCI), parts of the National Institutes of Health (NIH), have teamed with Group Health Cooperative in Seattle and Henry Ford Health System in Detroit to launch a study to investigate the interest level of healthy, young adults in receiving genetic testing for eight common conditions. Called the Multiplex Initiative, the study will also look at how people who decide to take the tests will interpret and use the results in making their own health care decisions in the future.

Significant Events in Population Genomics 2003 - 2007

- Major advances in common disease genetics
- A new paradigm for data sharing
- Developing approaches to speed up functional genomics
- Early indications that genomics may be highly relevant to health disparities in common disease
- Appreciation of need for better assessment of E and GxE
- Discussion, but not much resolution, of the need for large scale population cohort studies in the U.S.
- Introduction of genetic risk factor assessment into mainstream medicine – too soon?

I not only use all the brains I have, but all I can borrow.

Woodrow Wilson