

# **Population Genomics at NHGRI: History and Need**

**Frontiers in Population Genomics Workshop**

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

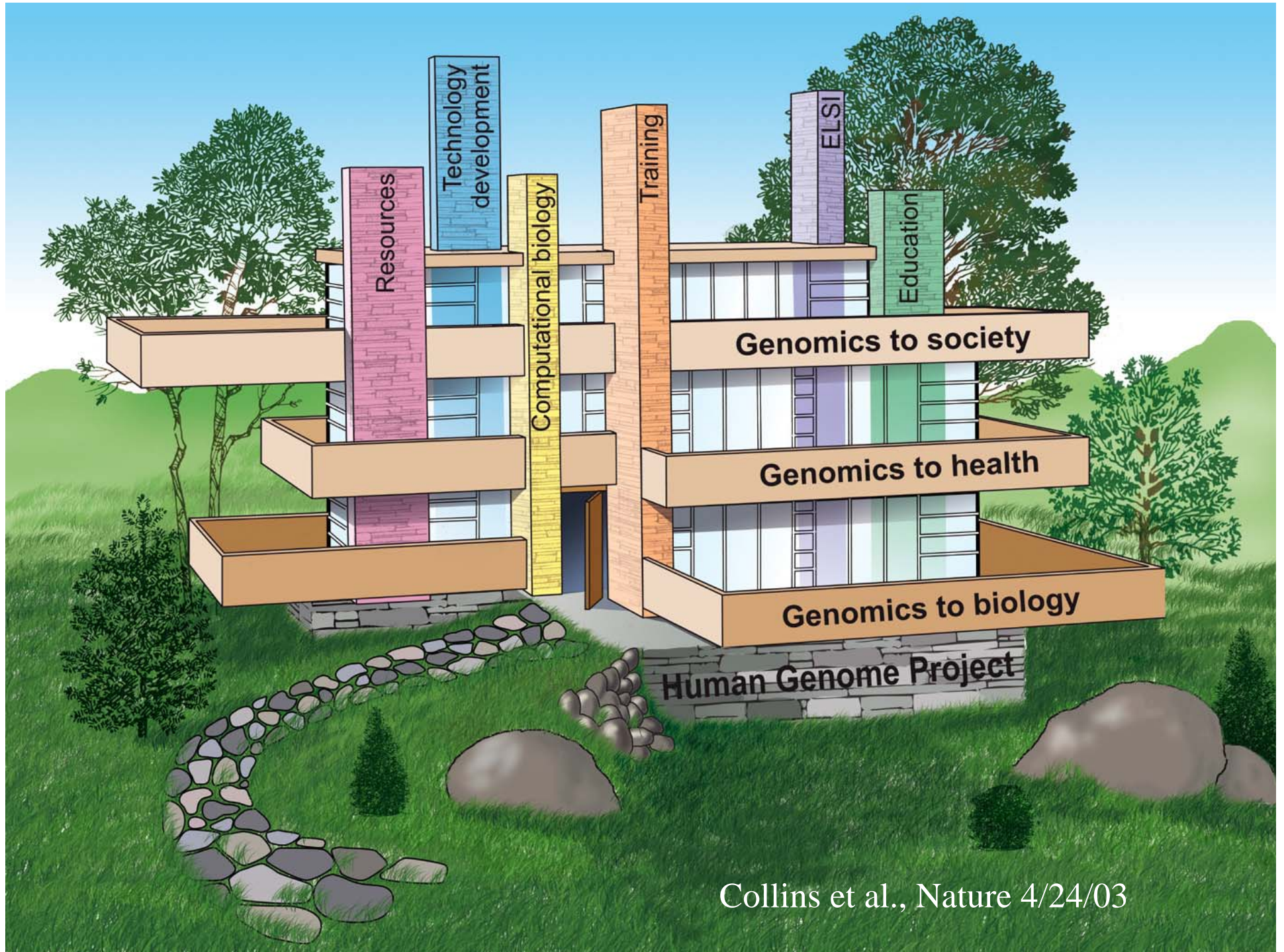
**December 18, 2007**



# SALLY FORTH



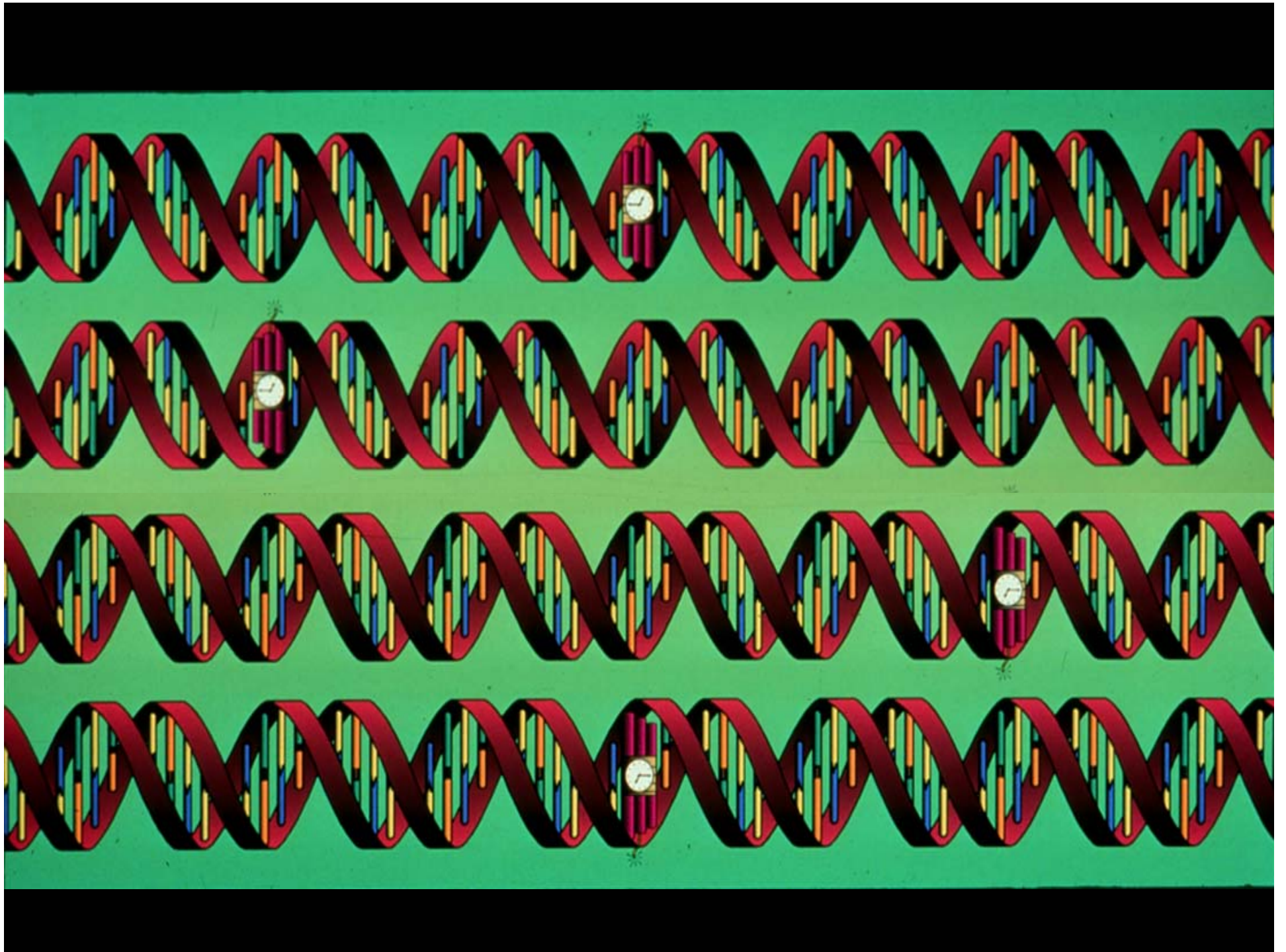
©2007 by King Features Syndicate, Inc. World rights reserved.



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

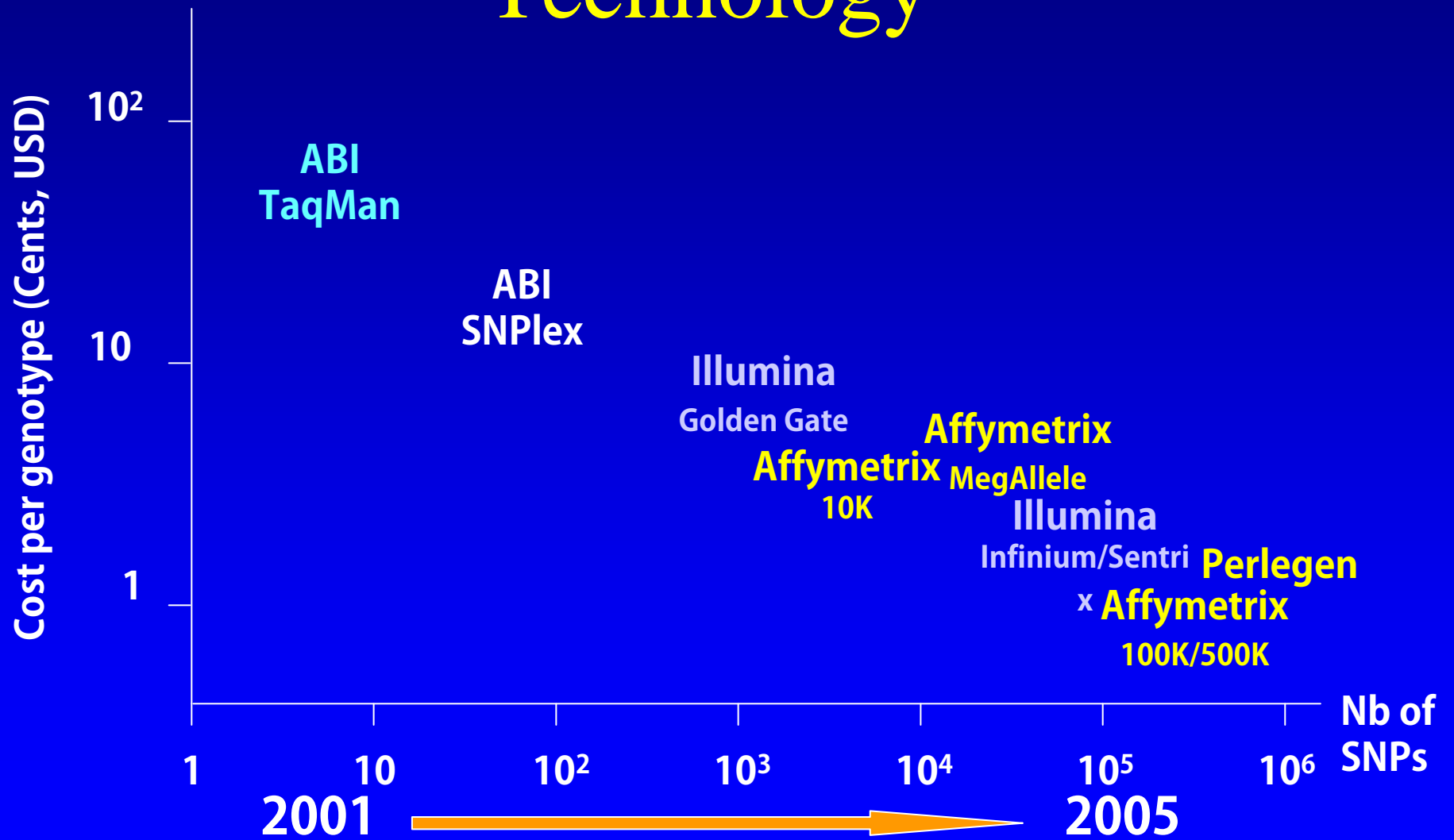
# nature



## THE HAPMAP PROJECT

Chapter and verse on  
human genetic variation

# Progress in Genotyping Technology



Courtesy S. Chanock, NCI



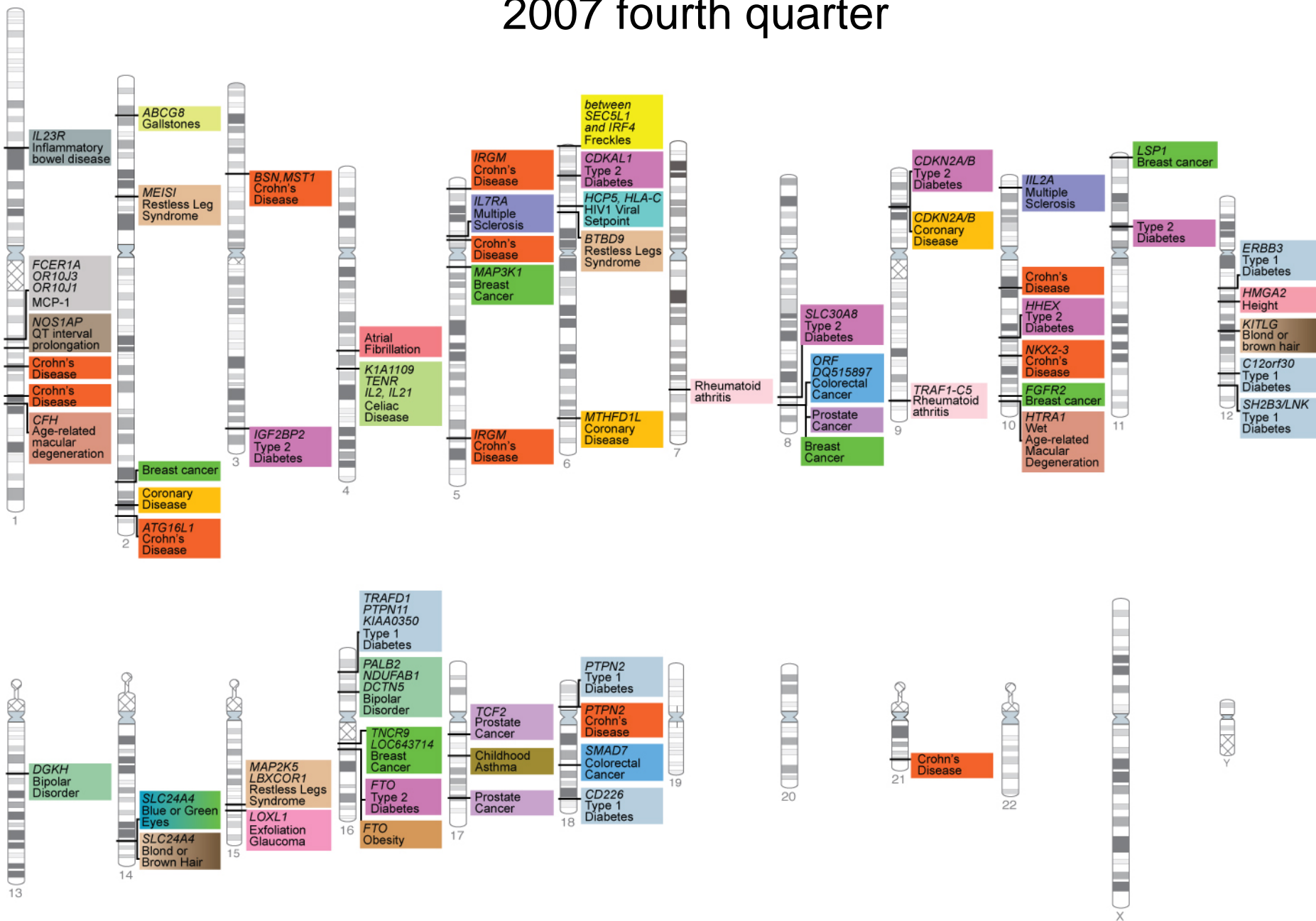
# **Genome Wide Association Approach to Common Disease: The View from 2007**

- **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 assemblies >1kb (preferably daily)
- Immediate submission of finished annotated sequence

- ~~aim~~ Aim to have all sequence freely available for both research and development, in order to maximise its benefit to society. and in the public domain

## 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.**



NCBI Search WGA for all[sb]

My NCBI Welcome mmailman, [Sign Out]

Study	Sub-Studies	Variables	Documents	Participants	Type of Study	Status
Age-Related Eye Disease Study (AREDS)	-	182	23	600	case-control	
NINDS Parkinsonism Study		400				
NINDS Parkinsonism Study - Cases						

Use the slider to adjust the p-value < 10-e 0

**-log10(p-value)**  
N/A <2 2-3 3-4 4-5 5-6 6-7 >7

Chr 1 10M 20M 30M 40M 50M 60M 70M 80M 90M 100M 110M 120M 130M 140M 150M 160M 170M 180M 190M 200M 210M 220M 230M 240M

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Chromosome 1 From 194000000 To 196000000 P-value 0.001 Reset Go

Gene	Start	End	P-value	Other	Tested	
1053003	1	195140223	1.08e-06	2/88 0.314	0.3618597	NA Tested Traits
3790414	1	195186922	2.51e-06	22/88 0.111	0.1349457	NA Tested Traits
7531555	1	195195933	4.87e-06	23/88 0.108	0.1892842	NA Tested Traits
12731209	1	195213762	9.35e-06	20/88 0.045	0.06132307	NA Tested Traits
1759016	1	195219121	2.07e-06	11/88 0.225	0.4517761	NA Tested Traits
10922152	1	195229629	2.63e-17	7/88 0.297	0.3117582	NA Tested Traits
10922153	1	195245238	2.63e-17	6/88 0.297	0.3117582	NA Tested Traits
16630903	1	195247303	9.00e-06	0.000 0.307	0.2137303	NA Tested Traits

MapViewer  
 MapViewer  
 Gene  
 RefSeq Transcr  
 Unigene  
 Total SNP  
 STS  
 OMIM  
 Morbid/Disease  
 Ideogram



## NCBI WGA Document Age Related Eye Disease Study

### Chapter 7 EXAMINATION PROCEDURES

#### 7.1 INTRODUCTION

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.

#### 7.2 REFRACTION AND VISUAL ACUITY

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. Pnhole 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 Visual Acuity Equipment and Facilities

**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:

Lighthouse Low Vision Products  
 36-02 Northern Boulevard  
 Long Island, New York 11101



# **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**



**Sequencing interesting regions to find causative variants**



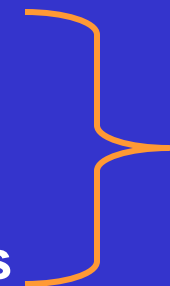
**Functional analysis**



**Translation**

- **Diagnostics**
- **Therapeutics**

**Data Analysis**



# **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)**

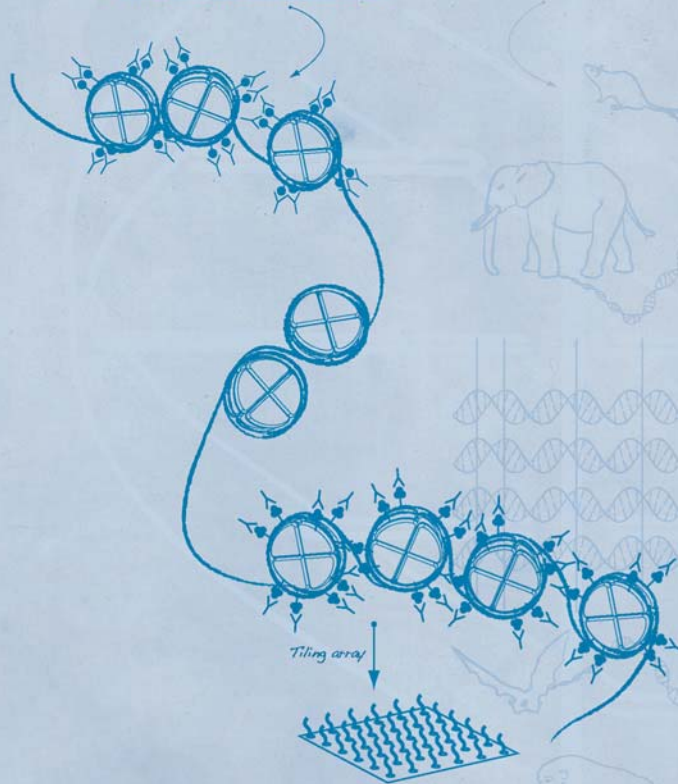
14 June 2007 | www.nature.com/nature | \$10

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

# nature

*Histone-modification chromatin IP*

*Comparative-sequence alignment*



**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



# **Significant Events in Population Genomics 2003 - 2007**

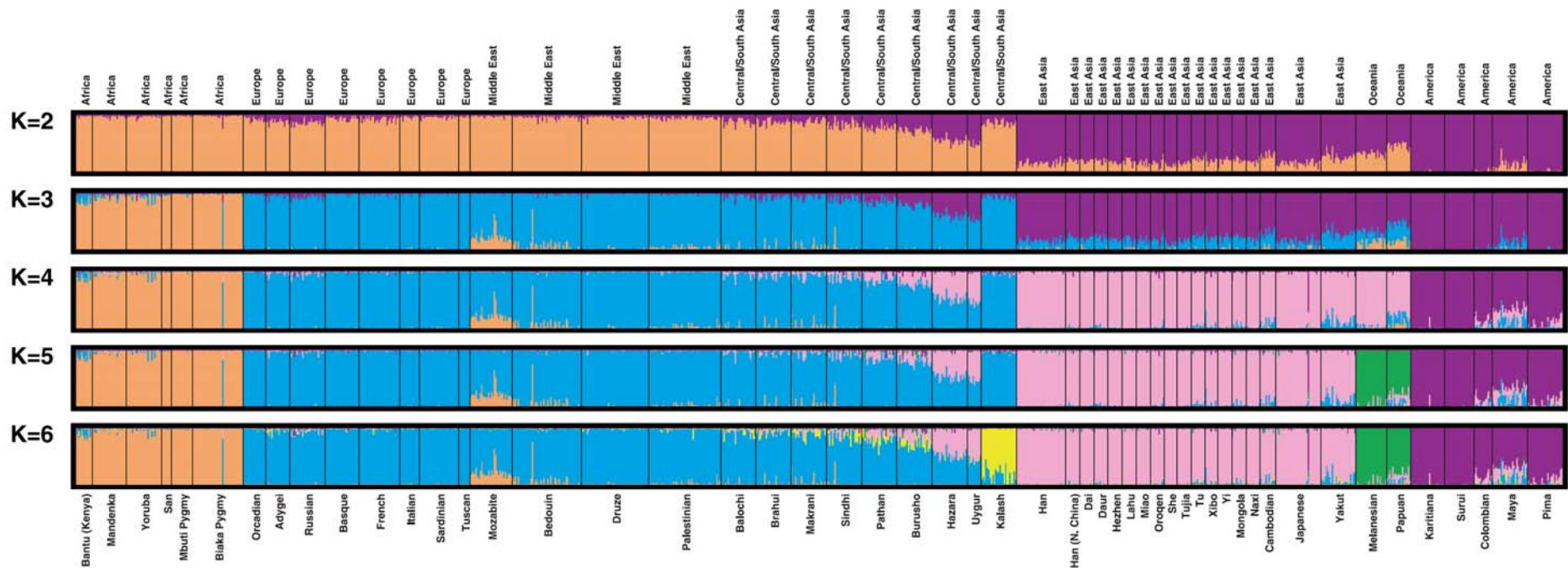
- **Major advances in common disease genetics**
- **A new paradigm for data sharing**
- **Developing approaches to speed up functional genomics**
- **Early indications that genomics may be highly relevant to health disparities in common disease**

# Genetic Structure of Human Populations

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

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<sup>1,12</sup>, Patrick Sulem<sup>1,12</sup>, Julius Gudmundsson<sup>1,12</sup>, Agnar Helgason<sup>1</sup>, Adam Baker<sup>1</sup>,

E  
M  
E  
C  
C  
J  
E  
C  
D  
8

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

Julius Gudmundsson<sup>1,12</sup>, Patrick Sulem<sup>1,12</sup>, Agnar Helgason<sup>1</sup>, Laufey T Amundadottir<sup>1,12</sup>

Multiple regions within 8q24 independently affect risk for prostate cancer **Prostate Cancer**

Christopher A Haiman<sup>1</sup>, Nick Patterson<sup>2</sup>, Matthew L Freedman<sup>2,3</sup>, Simon R Myers<sup>2</sup>, Malcolm C Pike<sup>1</sup>,

K Ali

G Ste

W Da

A Ka

Br

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

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

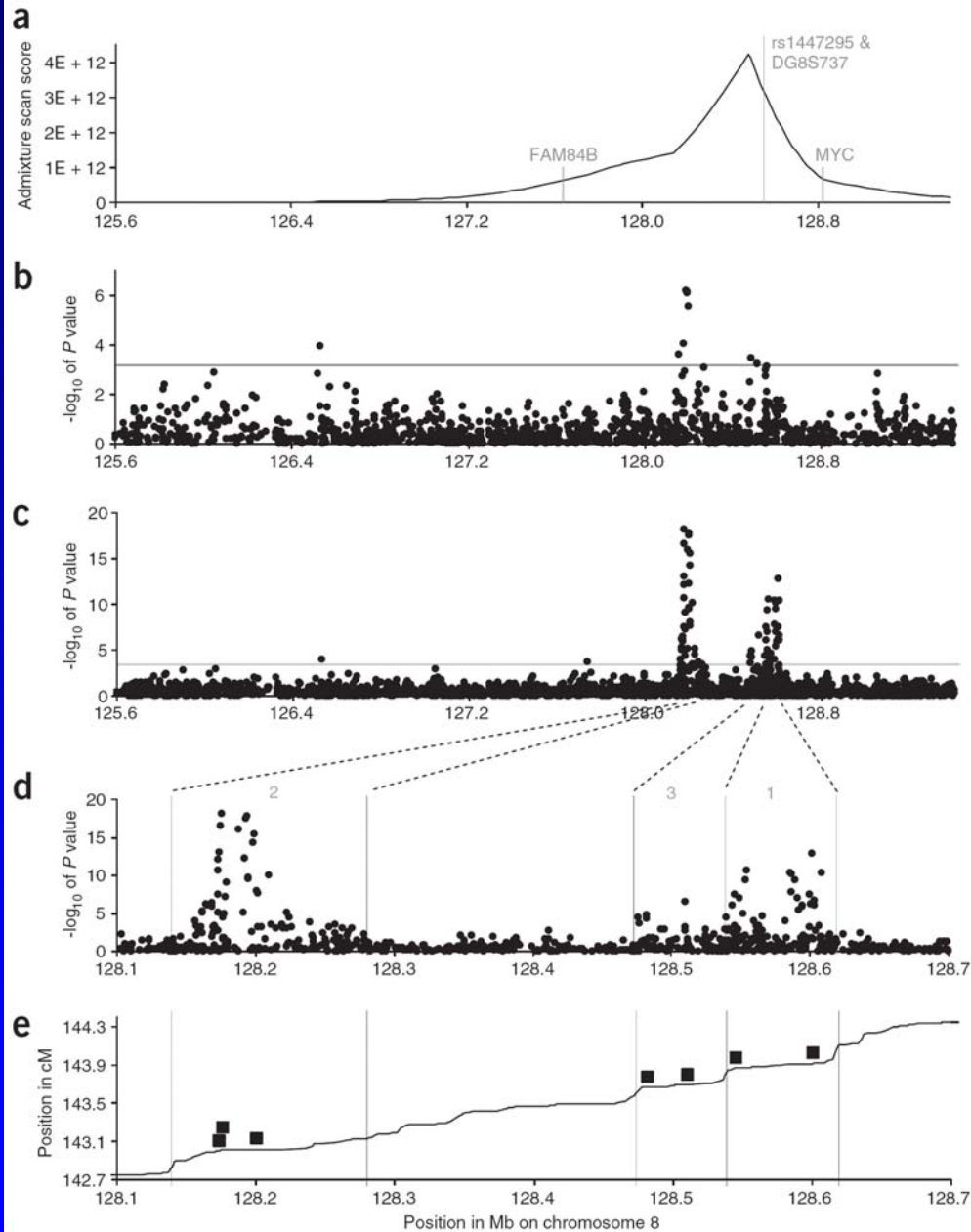
## Association of DG8S737 “Allele – 8” with Prostate Cancer

---

	Allele Frequency		OR	P-value
	Cases	Controls		
Iceland	0.131	0.078	1.77	$2 \times 10^{-8}$
Sweden	0.101	0.079	1.38	$4 \times 10^{-3}$
Chicago (European)	0.082	0.041	2.10	$3 \times 10^{-3}$
Michigan (African-American)	0.234	0.161	1.60	$2 \times 10^{-3}$

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

## GENETICS PROGRAM



Identify genetic variants

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

**GXE**

# Gene Environment Association (GENEVA) Studies

<b>Principal Investigator</b>	<b>Institution</b>	<b>Primary Outcome</b>
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**

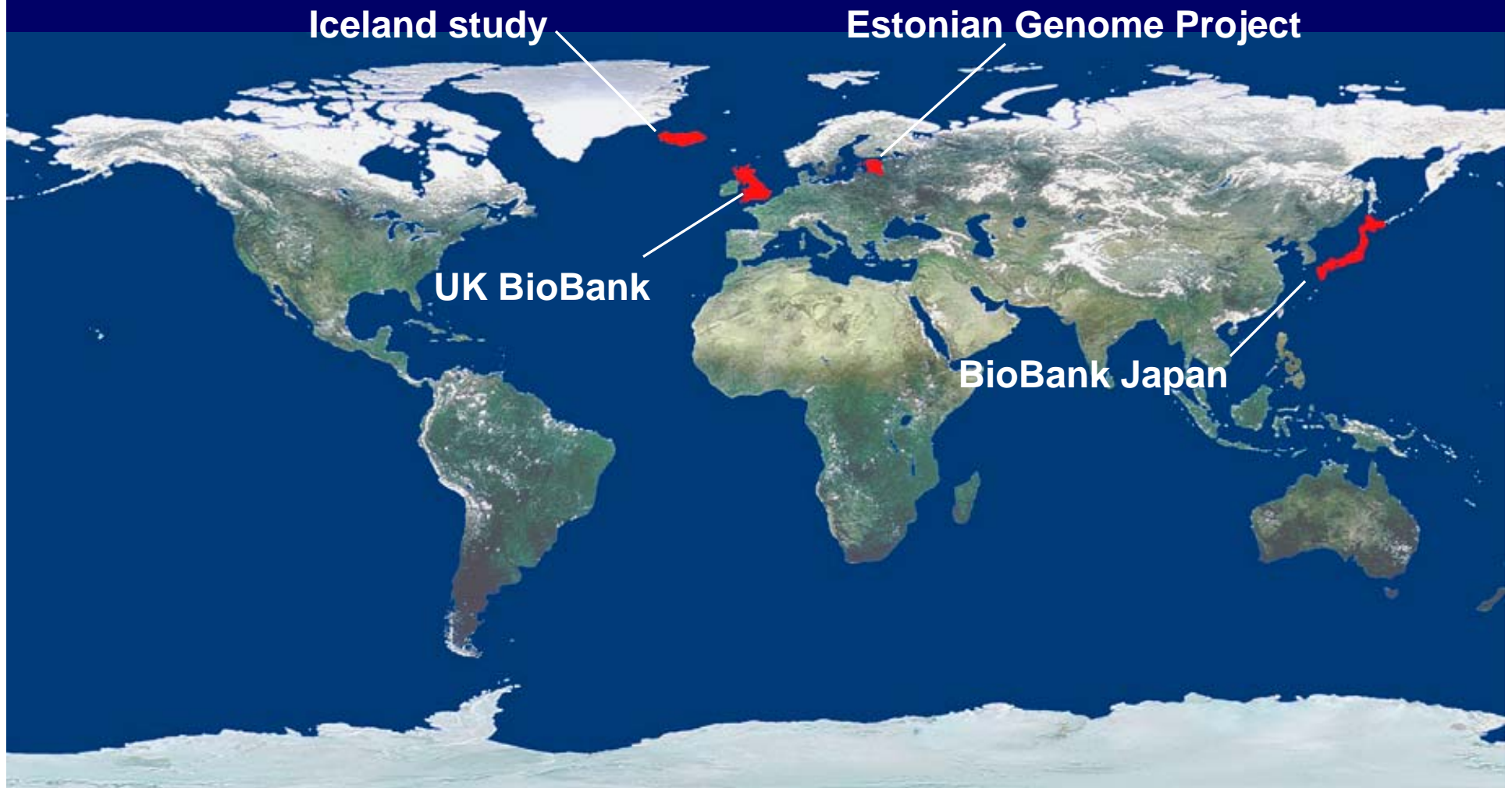
# The case for a US prospective cohort study of genes and environment

**Francis S. Collins**

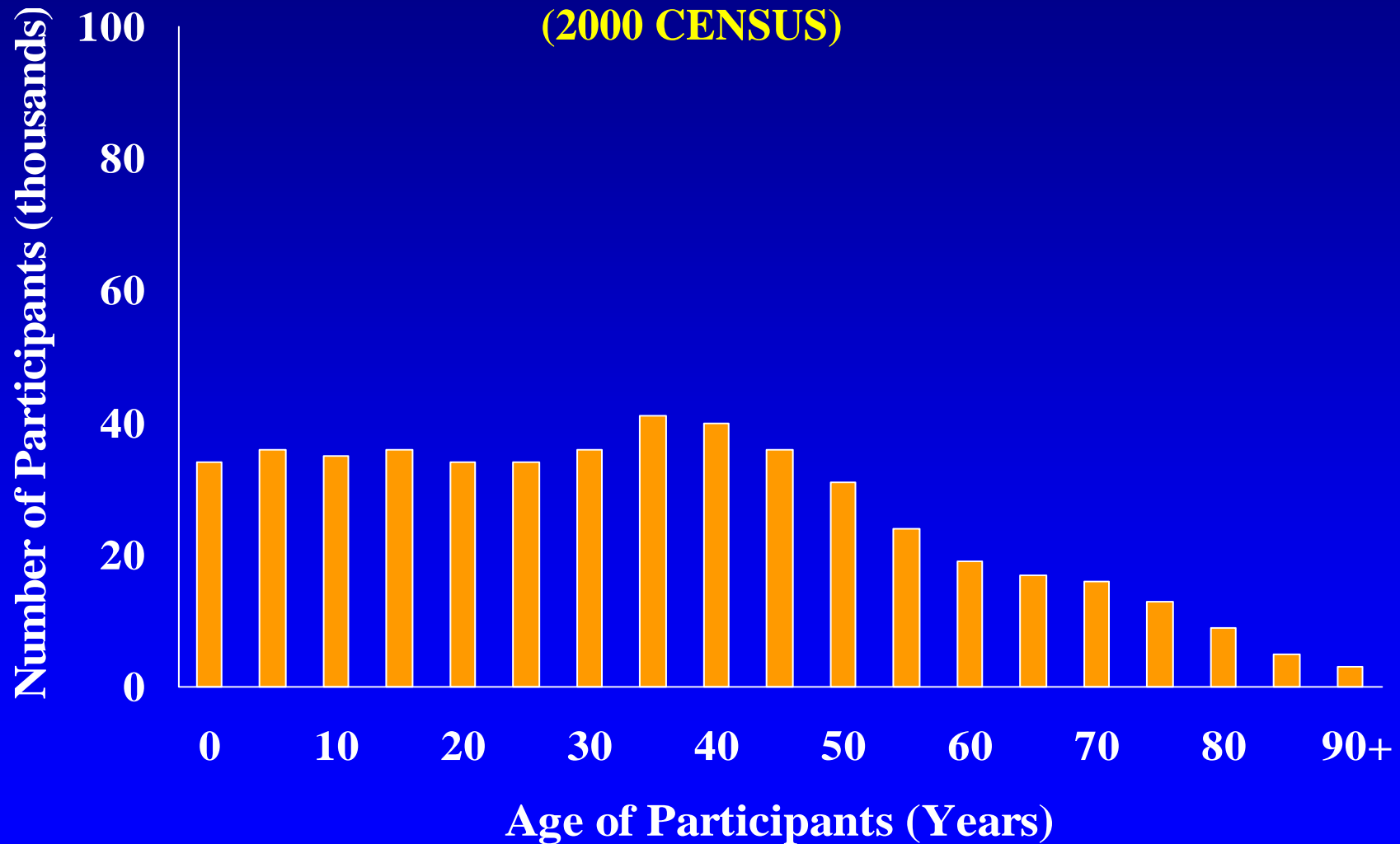
*National Human Genome Research Institute, National Institutes of Health, Building 31, Room 4B09, MSC 2152, 31 Center Drive, Bethesda, Maryland 20892-2152, USA (e-mail: fc23a@nih.gov)*

---

**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**

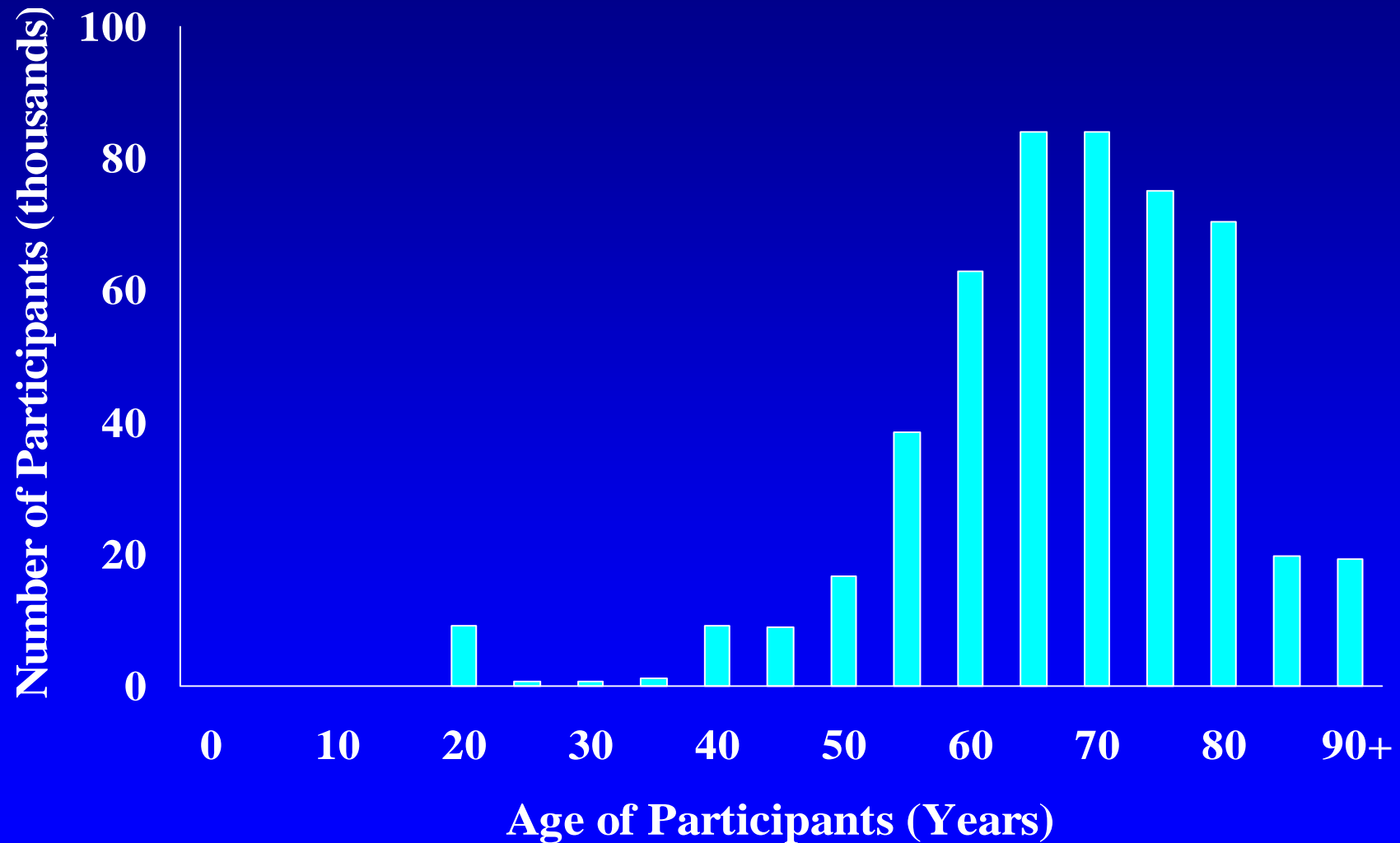


# ESTIMATED AGE DISTRIBUTION OF REPRESENTATIVE US COHORT (2000 CENSUS)





# 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**





Microsoft Internet Explorer browser window displaying the Navigenics website. The address bar shows <http://www.navigenics.com/>. The website header features the Navigenics logo and navigation links: [About](#), [Leadership](#), [Policies](#), and [Contact](#).

The main content area includes a video player with a thumbnail of a man sitting on a couch. The video title is "My Genes. My Health. My Life. My Guide." and there is a "Play Video" button. Below the video, the text reads: "Your genes offer a road map to optimal health".

To the right of the video, the "Welcome to Navigenics" section contains the following text:

**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.

The footer of the website includes a navigation menu: [Home](#) | [About](#) | [Leadership](#) | [Policies](#) | [Contact](#) and a copyright notice: Copyright © 2007 Navigenics, Inc. All rights reserved.

The Windows taskbar at the bottom shows the Start button, several open applications (Feedreader, Microsoft Office, Slides-final, Microsoft Office, Internet Explorer), and the system tray with the time 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**