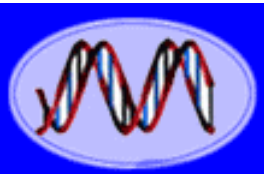
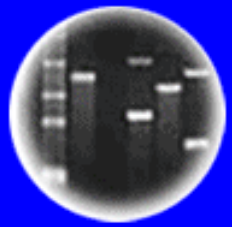


DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

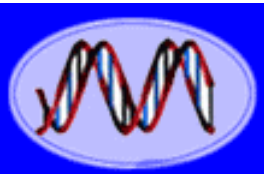
HC70A & SAS70A Winter 2011
Genetic Engineering in Medicine, Agriculture, and Law

Professors Bob Goldberg & John Harada

Lecture 6
Your Personal Genome &
Tracing Your Ancestry

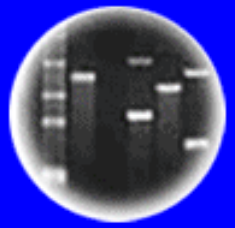
UCLA

UCDAVIS
UNIVERSITY OF CALIFORNIA



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

Last Lecture

1. Two Genomes in a Cell!
2. What is the Mitochondrial Genome and How is it Inherited?
3. How Can Mitochondrial Genome Markers Be used To Trace Our Ancestry Back to Eve?
4. What Are the Characteristics of the Human Genome?
5. The Human Genome Project
6. How Has DNA Sequencing Output and Costs Changed Since the Start of the Human Genome Project?
7. High Throughput DNA Sequencing & Genome Annotation

PERSPECTIVE

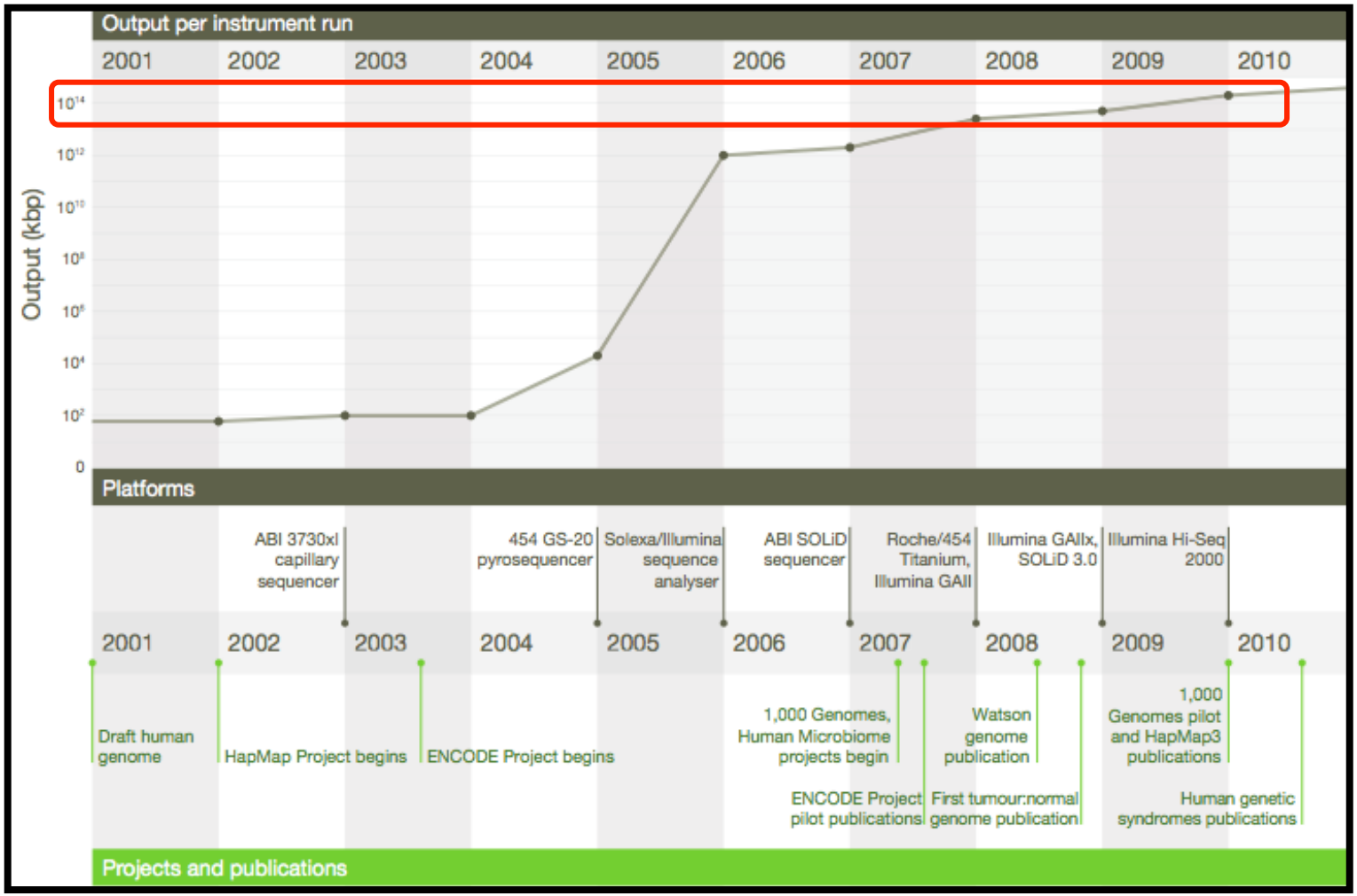
doi:10.1038/nature09796

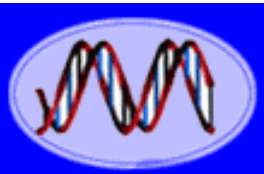
A decade's perspective on DNA sequencing technology

Nature, February 10, 2011

Elaine R. Mardis[1]

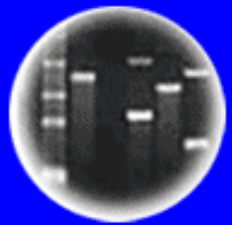
10^{14} kbp





DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

A New Comparative Genomics Field Has Emerged Allowing the Comparison of Entire Genomes!

TABLE 17.1

Representative Sequenced Genomes

ORGANISM	HAPLOID GENOME SIZE (Mb)	NUMBER OF GENES	PROTEIN-CODING SEQUENCE
Bacteria			
M. genitalium	0.58	485	88%
H. influenzae	1.8	1,738	89%
E. coli	4.6	4,377	88%
Yeasts			
S. cerevisiae	12.5	5,770	70%
S. pombe	12.5	4,929	60%
Plants			
A. thaliana	115	28,000	25%
Rice	390	37,544	12%
Animals			
C. elegans	100	19,427	25%
D. melanogaster	123	13,379	13%
Pufferfish	342	27,918	10%
Chicken	1,130	25,000	3%
Human	3,300	24,000	1.2%

Mb = millions of base pairs

Learning About "Life" By Peering Into Whole Genomes

TABLE 17.2

Comparison of the Genomes of *E. coli* and Yeast

	E. COLI	YEAST
Genome length (base pairs)	4,640,000	12,068,000
Number of protein-coding genes	4,290	5,770
Proteins with roles in:		
Metabolism	650	650
Energy production/storage	240	175
Membrane transport	280	250
DNA replication/repair/ recombination	120	175
Transcription	230	400
Translation	180	350
Protein targeting/secretion	35	430
Cell structure	180	250

TABLE 17.4

Arabidopsis Genes Unique to Plants

FUNCTION	NUMBER OF GENES
Cell wall and growth	42
Water channels	300
Photosynthesis	139
Defense and metabolism	94

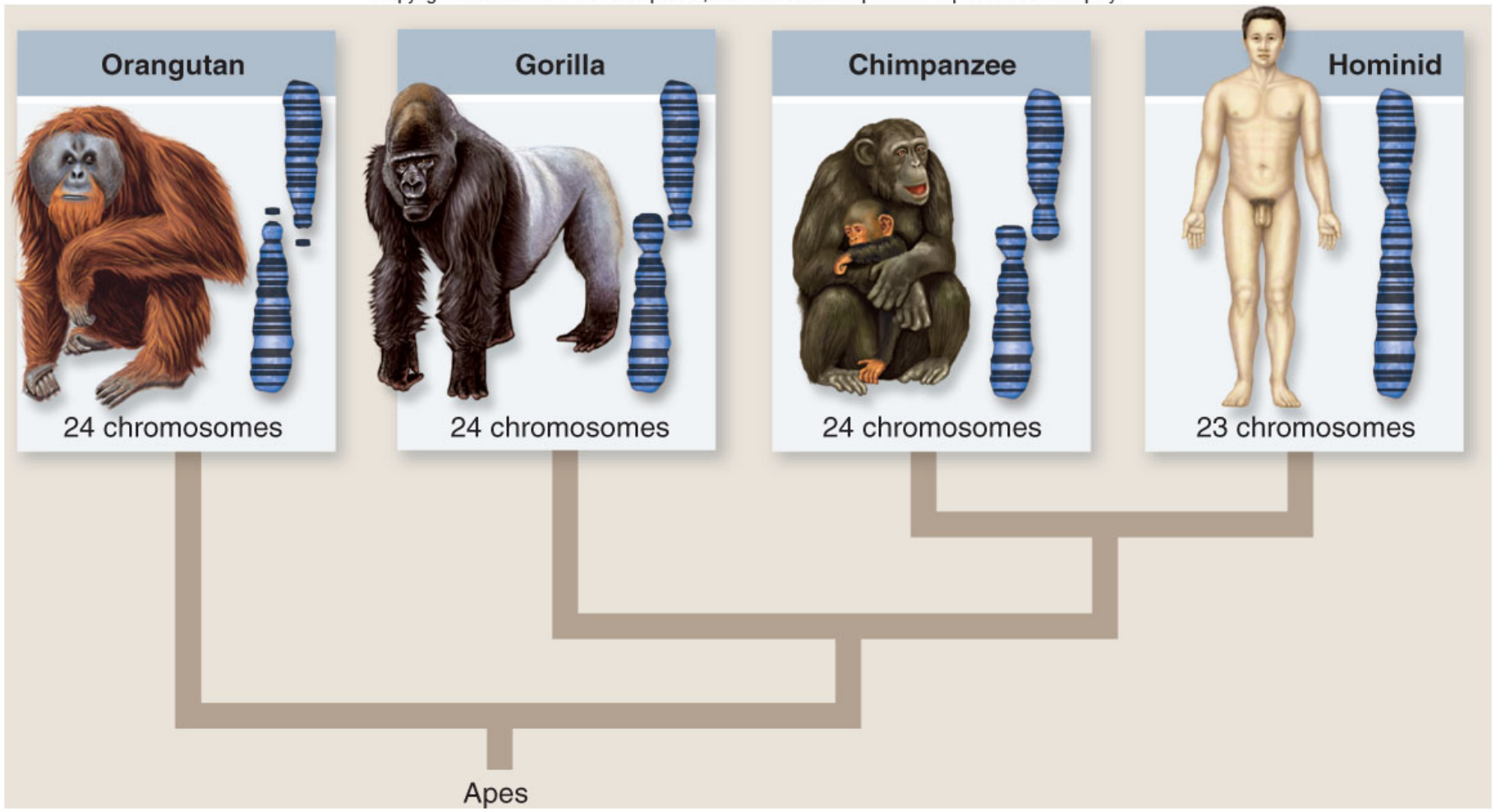
Many Mammalian Genomes Have Been Sequenced And More Are Being Sequenced

Human	Rabbit
Mouse	Rat
Dog	Ground Squirrel
Cow	Tree Shrew
Guinea Pig	Dolphin
Sloth	Chimpanzee
Armadillo	Gorilla
Kangaroo Rat	Orangutan
Horse	Rhesus Monkey
Cat	Wallaby

+ 1,000 Individual Human Genomes Including James Watson
Because of Major Breakthroughs in Sequencing Technology

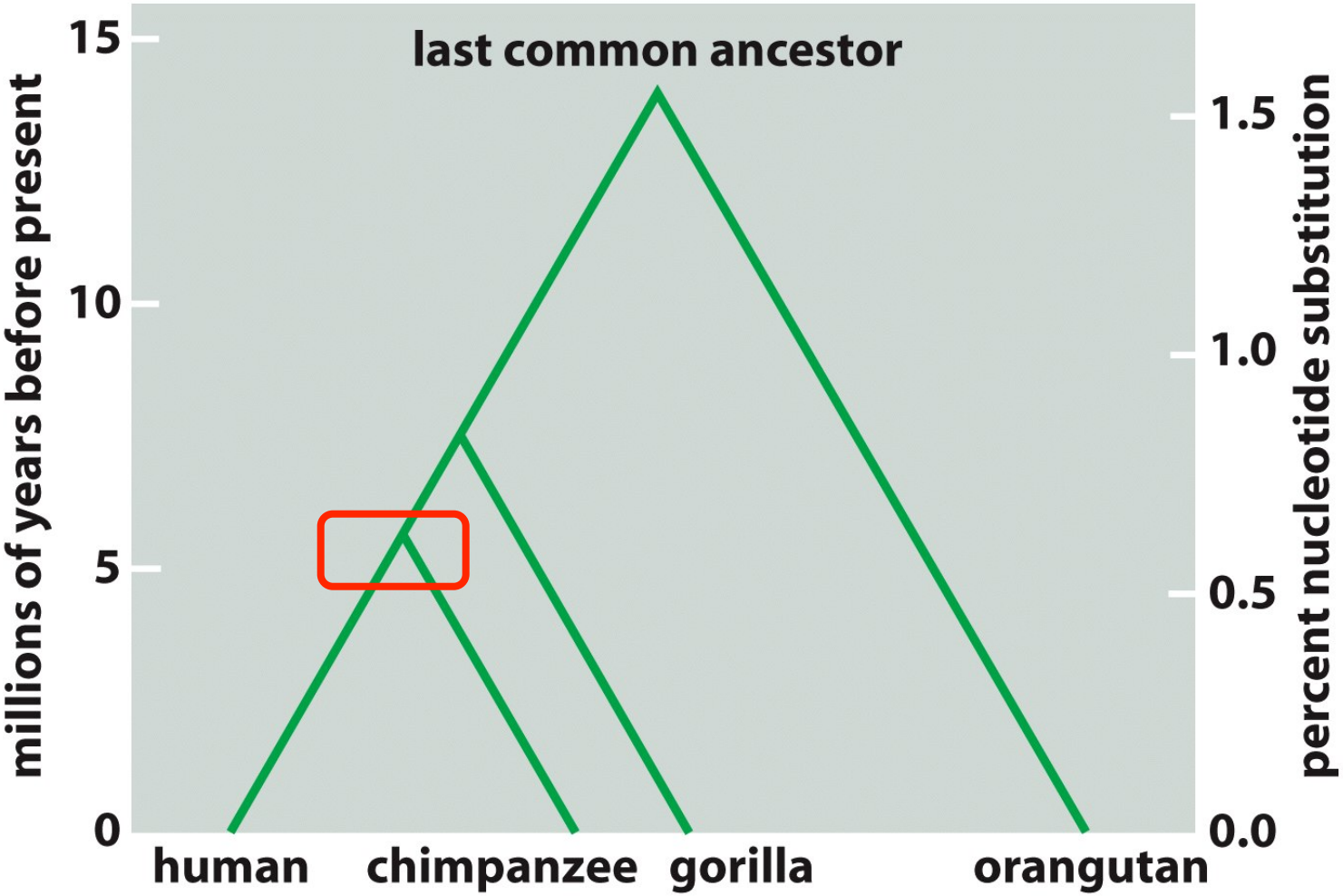
2011

Comparison of Mammalian Genomes Attempts To Determine "What Makes a Man, a Man and a Mouse a Mouse"



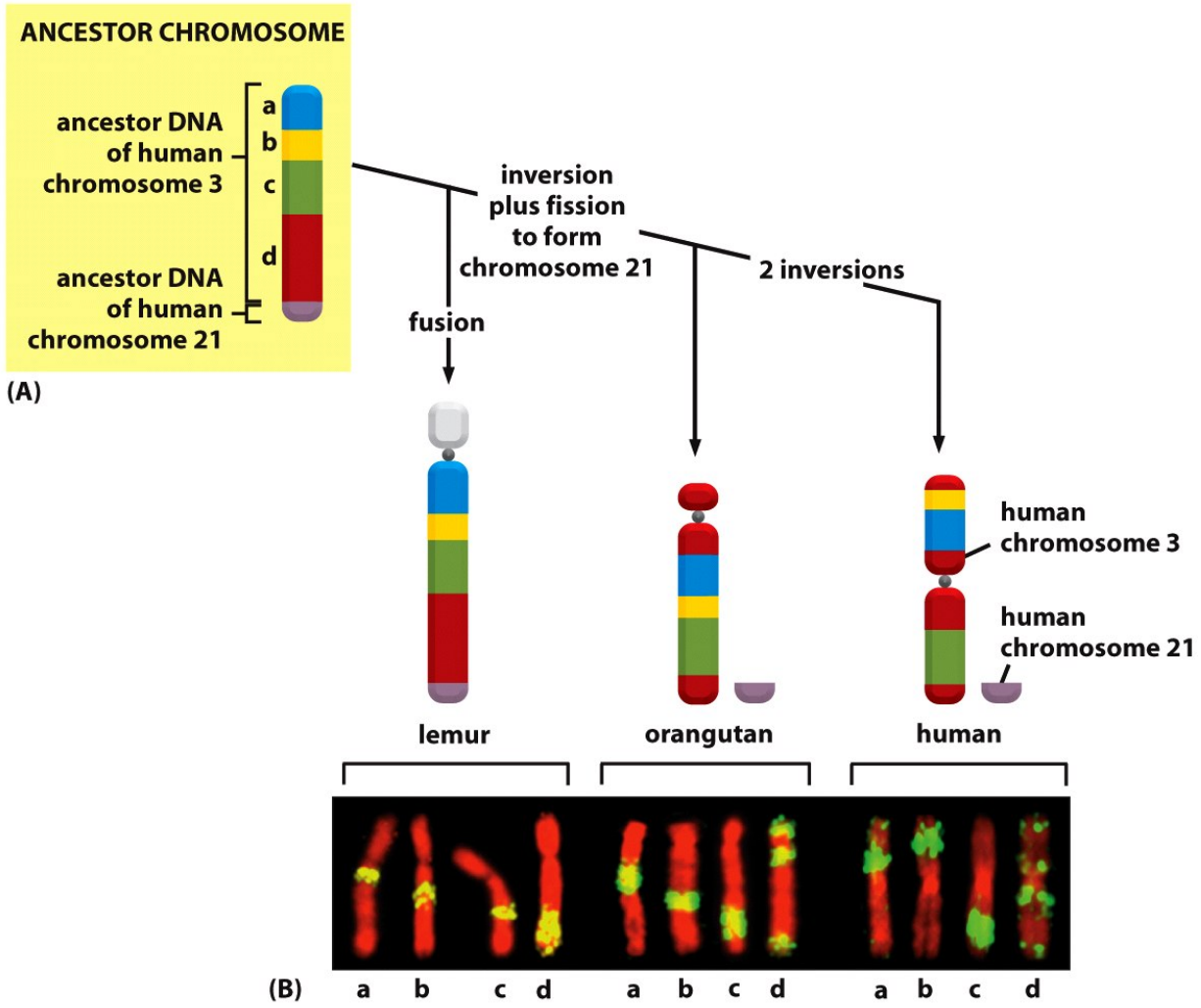
There is <1% Difference Between Human & Chimpanzee DNAs!

Comparison Between Primate Genomes

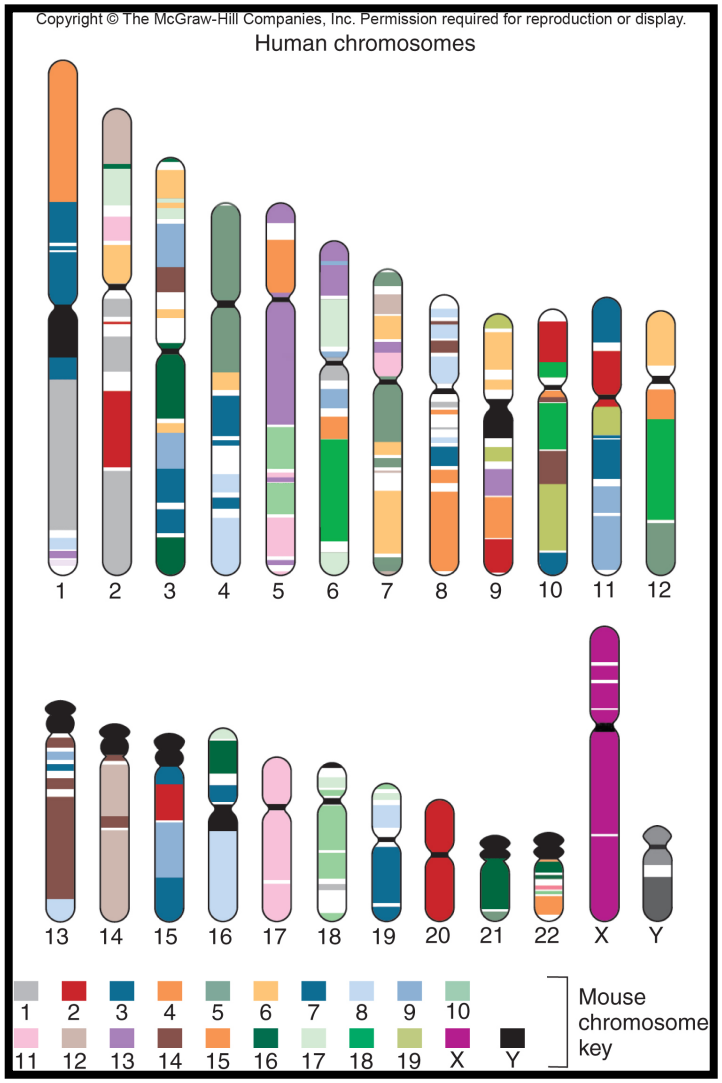


Note the Small Sequence Differences in These Genomes-
What Makes a "Human a Human?"

Comparative Genomics Can Uncover the Origin of Human Chromosomes and Relationship to Other Mammalian Chromosomes

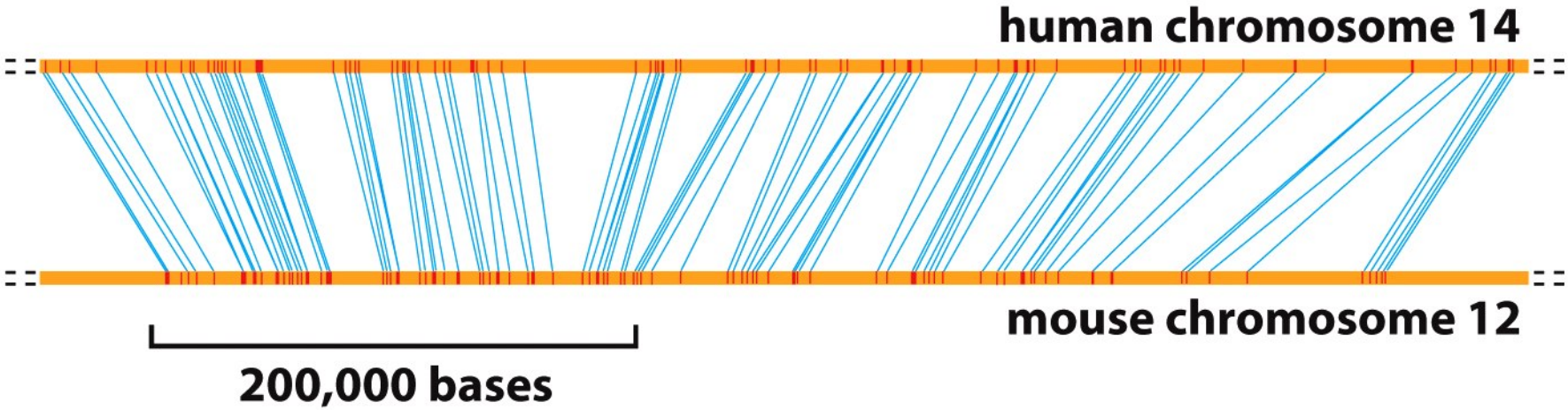


What Makes a Mouse a Mouse and a "Man a Man"



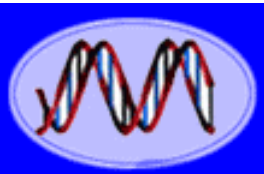
Note all of the Mouse Sequences in Human Chromosomes!

Comparative Genomics Can Align Related Genes in Two Different Genomes



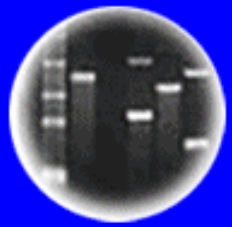
Note "Synteny," or Alignment, of Related Genes Between Human and Mouse Chromosome Regions

What Does This Say About Genome Evolution?



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria

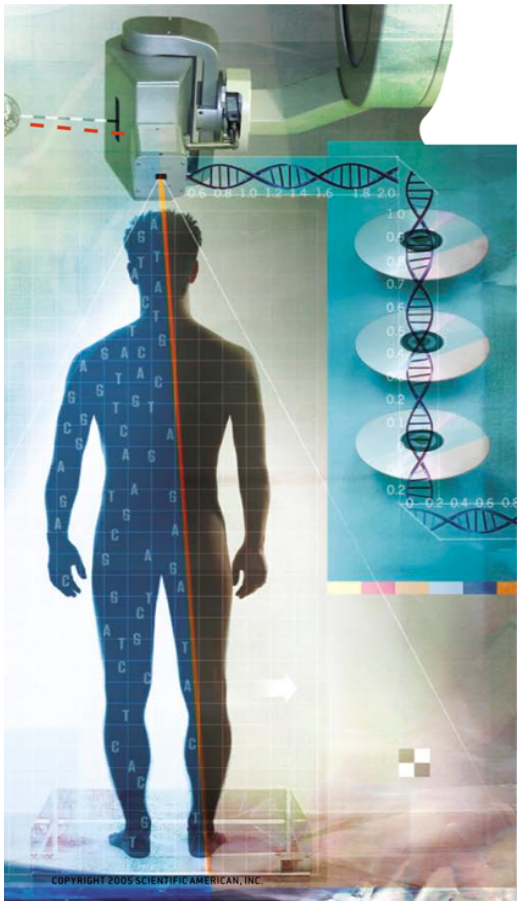


DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

The Personalized Genome

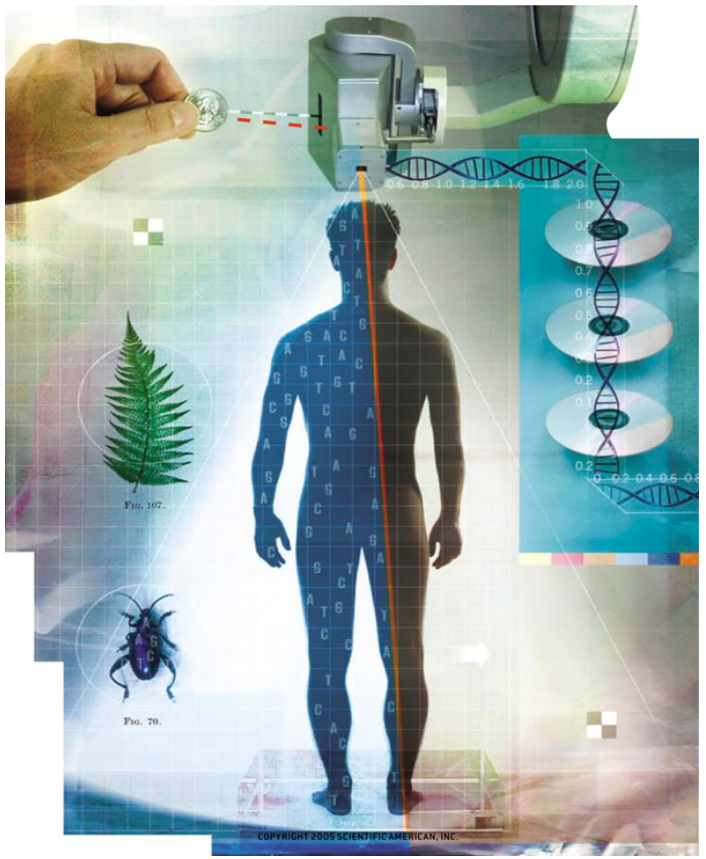


Ultimately-You <u>Are</u> What Is In Your Genome

The Ultimate Measure of Individuality
Personal Genome Sequence & Comparing Individual Human Genomes

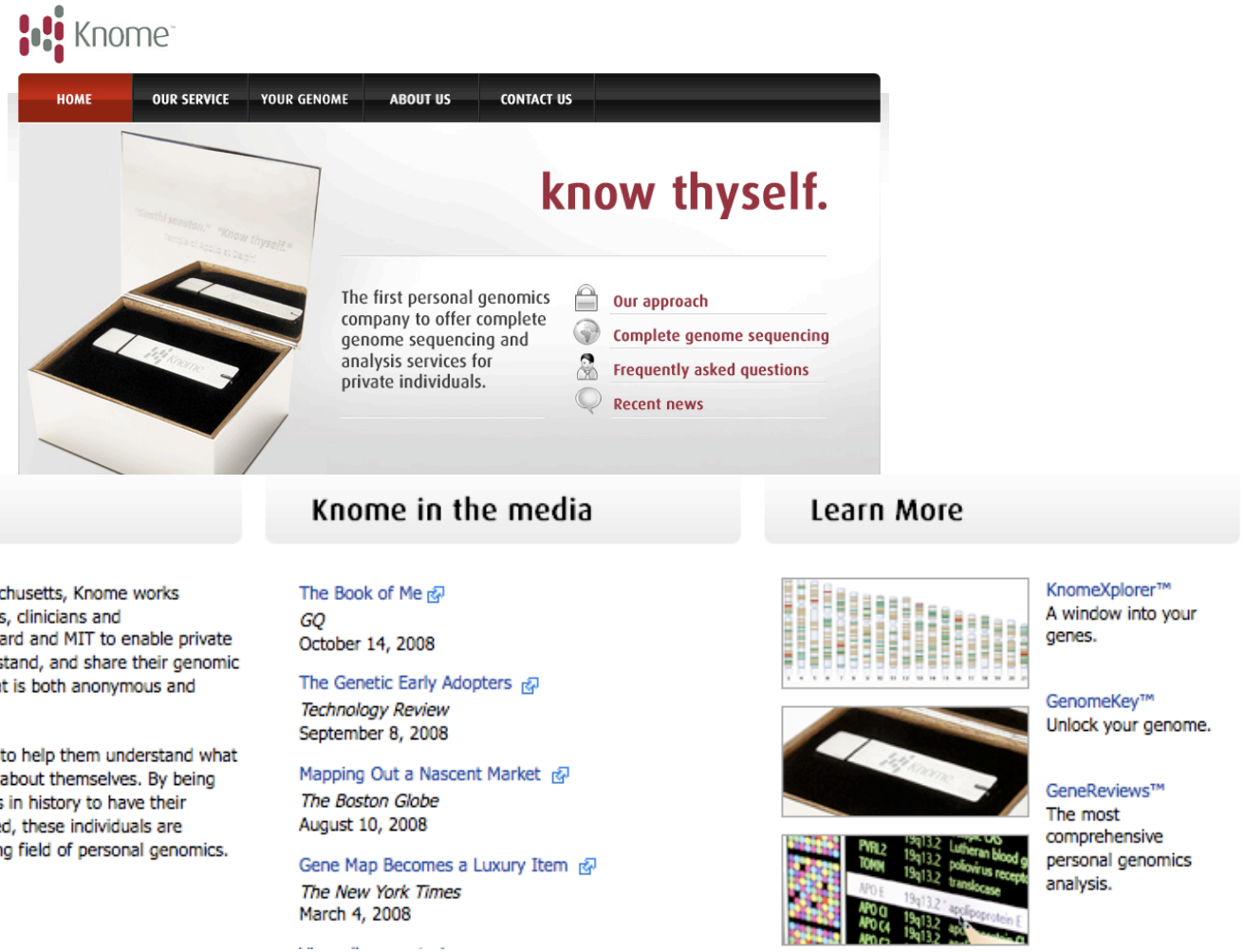
Genomes for ALL

Next-generation technologies that make
reading DNA fast, cheap and widely accessible
are coming in less than a decade.
Their potential to revolutionize research
and bring about the era of truly personalized medicine
means the time to start preparing is now



Find DNA Variability in
<u>All</u> Genes & Associate
with Specific Traits!

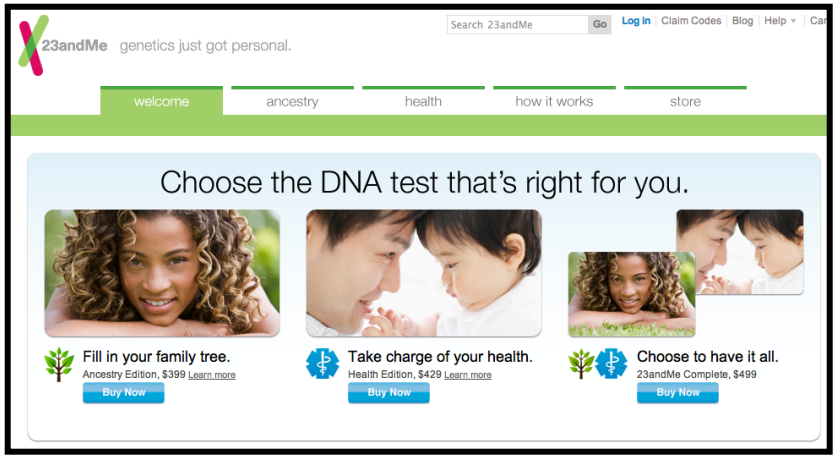
The Age of Personal Genomics Has Begun!

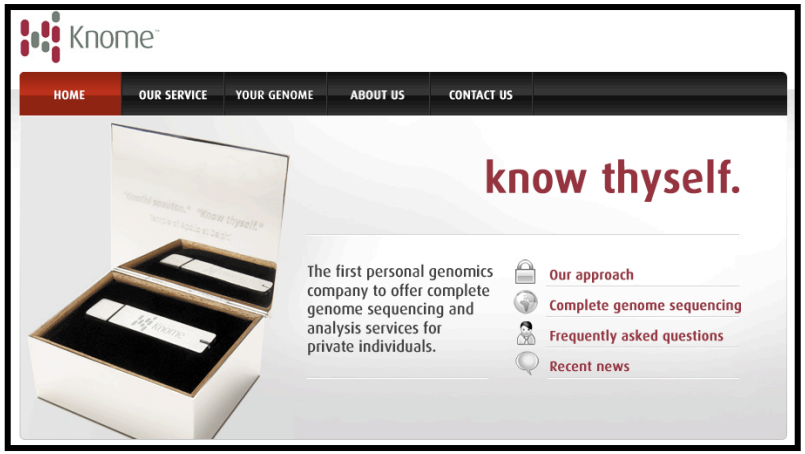


$50,000-Soon Down to $1,000

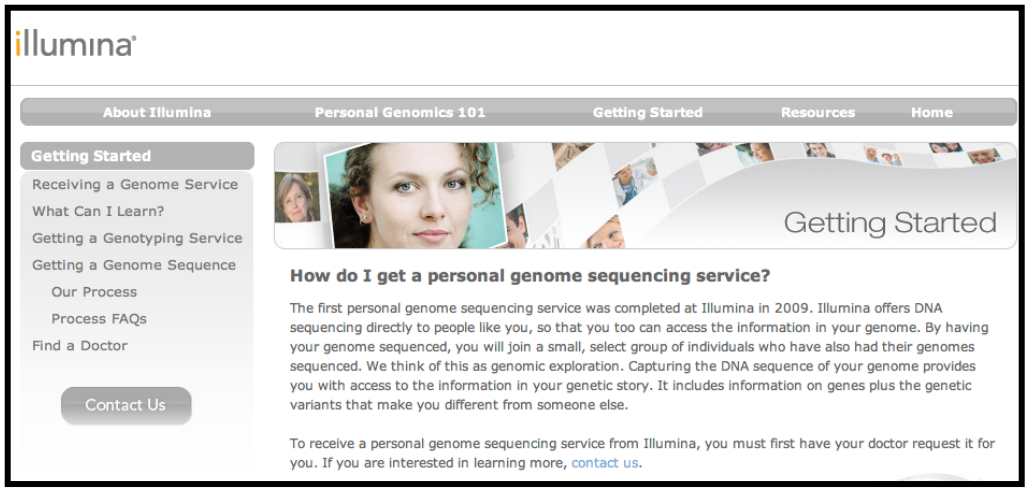
Problems?

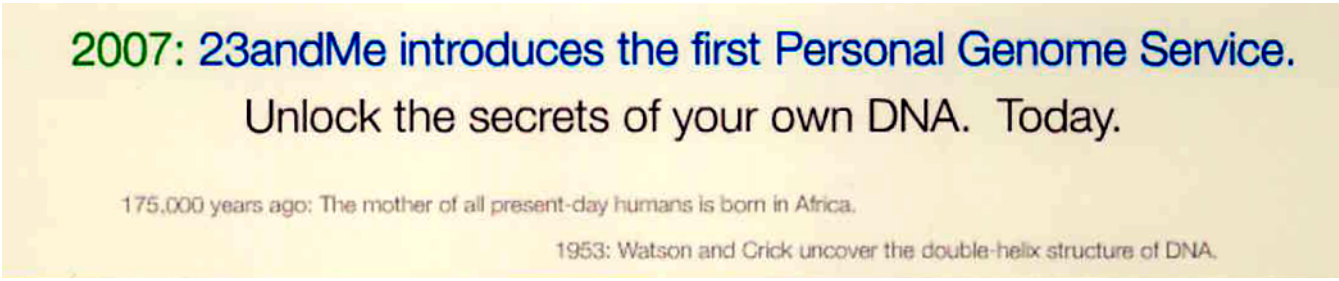
Personal Genome Sequencing Companies

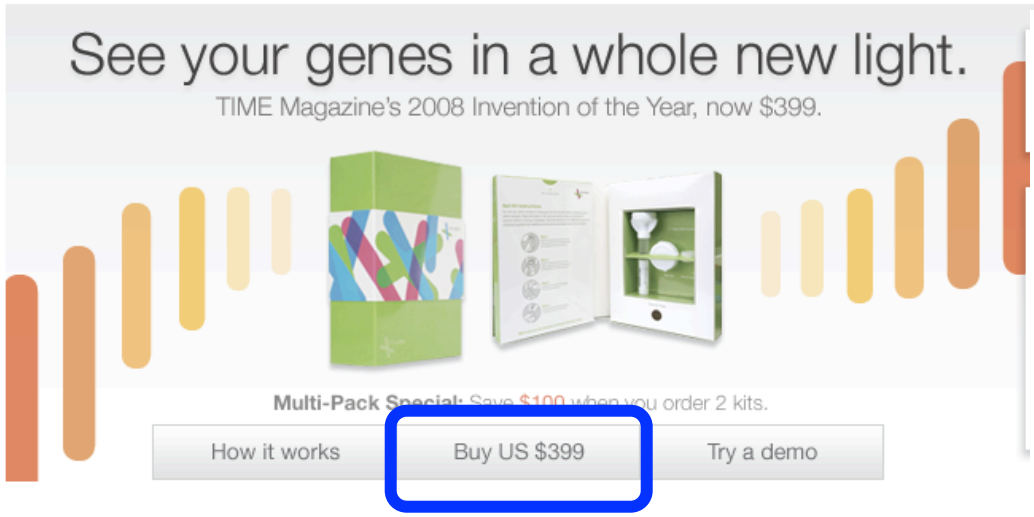


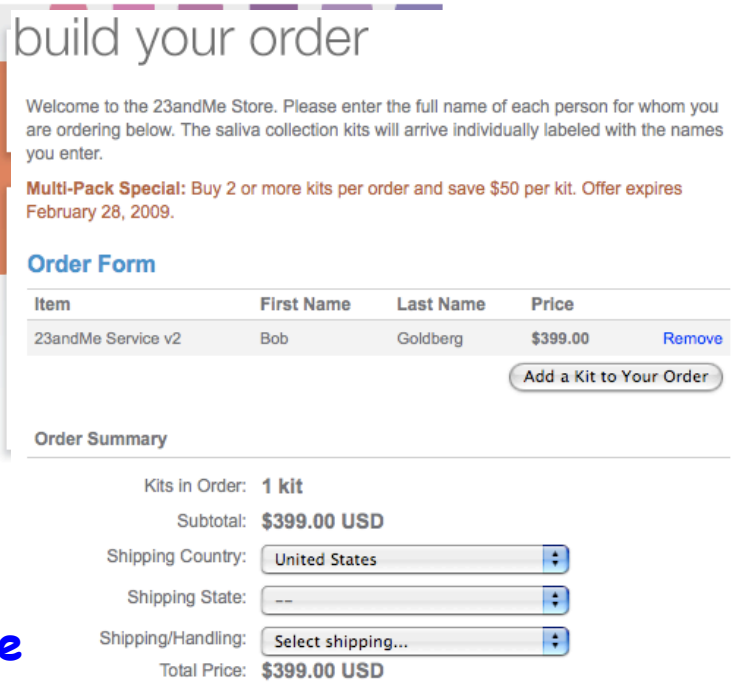


$1,000 Genome?



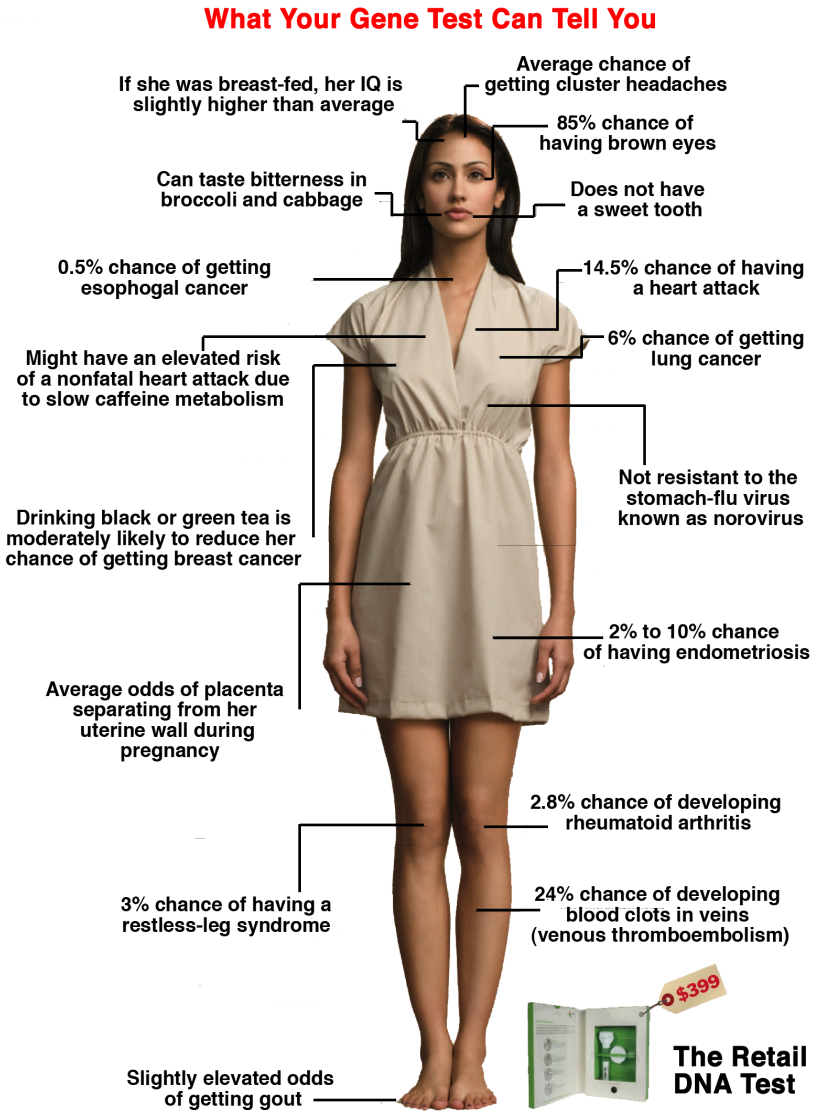






What Are the Problems With This Service and Approach to Personal Genomics?

Time Magazine 2008 - Invention of the Year
Your Personal Genome - 23andMe®



Invention Of the Year

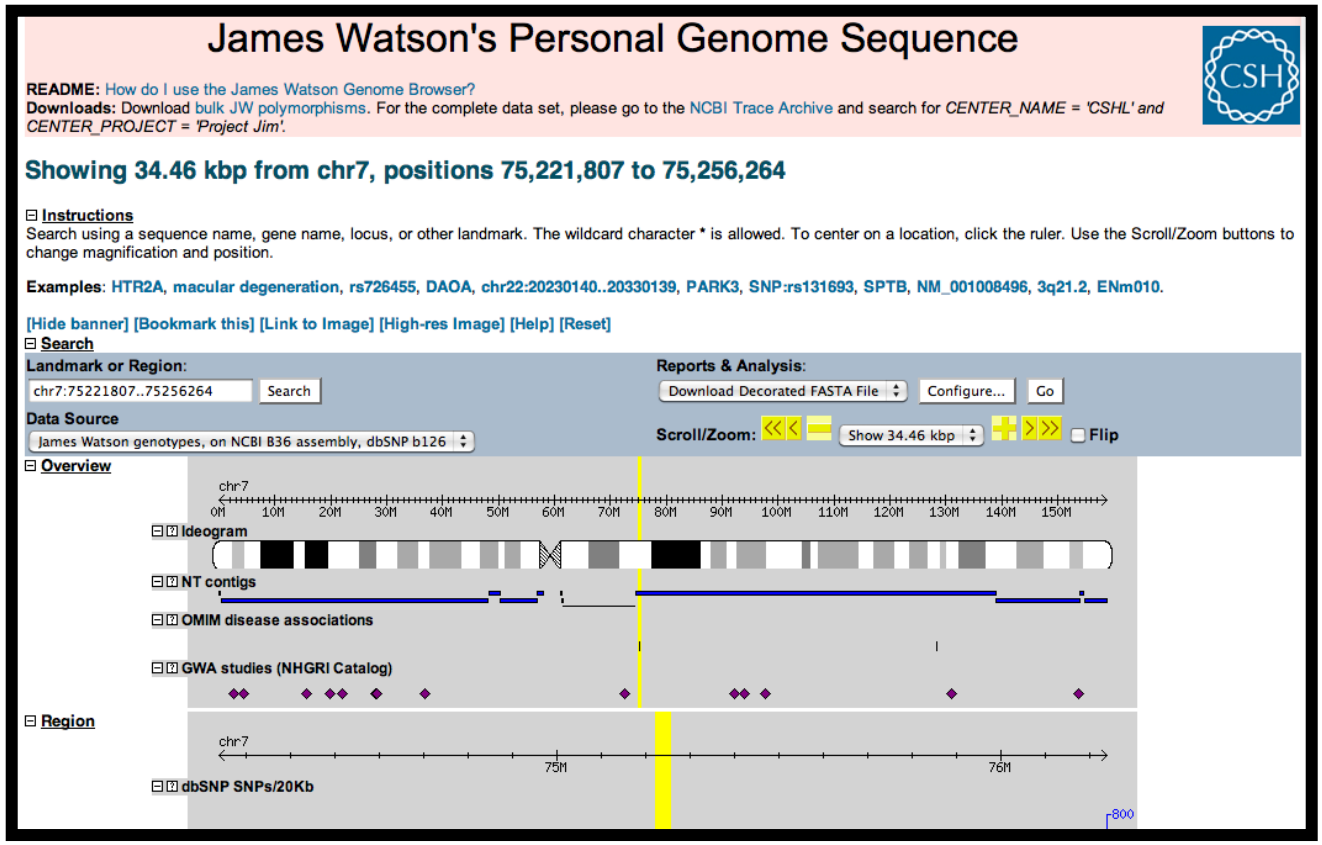
https://www.23andme.com/

And
Before Birth!!!

The Complete Genome of Individuals Can Now Be Decoded and Sequenced Very Inexpensively ($10,000)!!

Genome of DNA Pioneer Is Deciphered

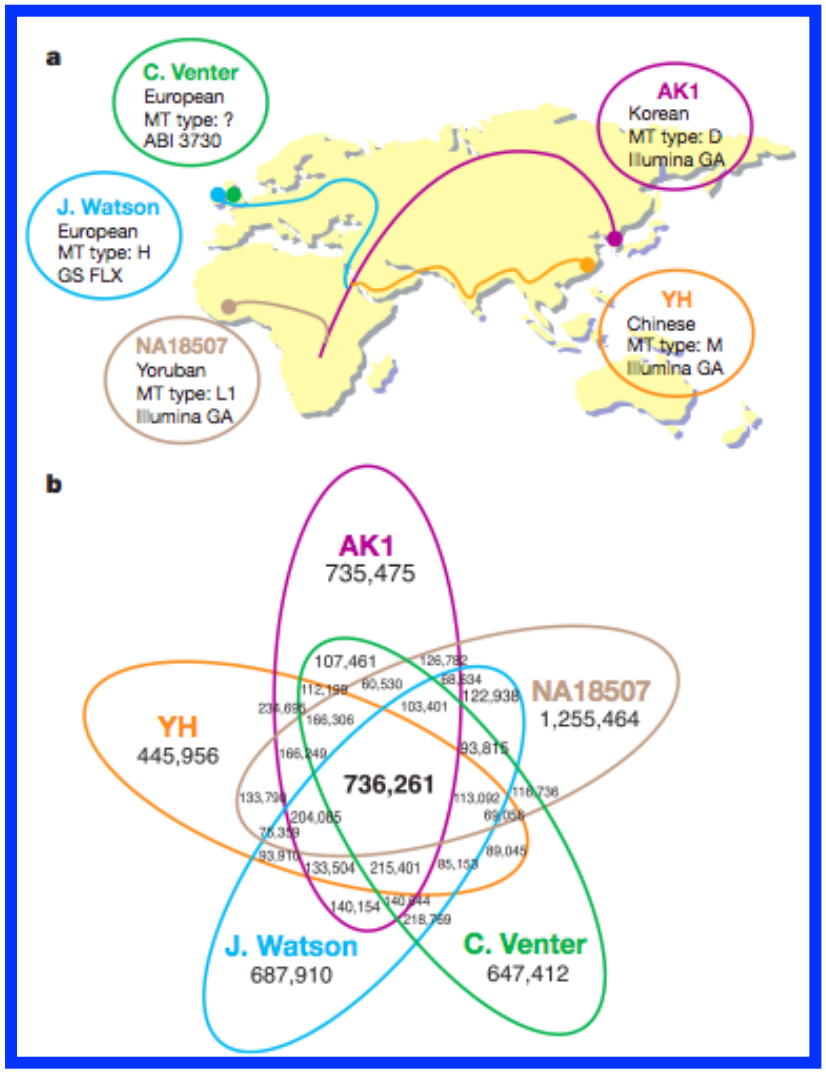
By NICHOLAS WADE
Published: May 31, 2007

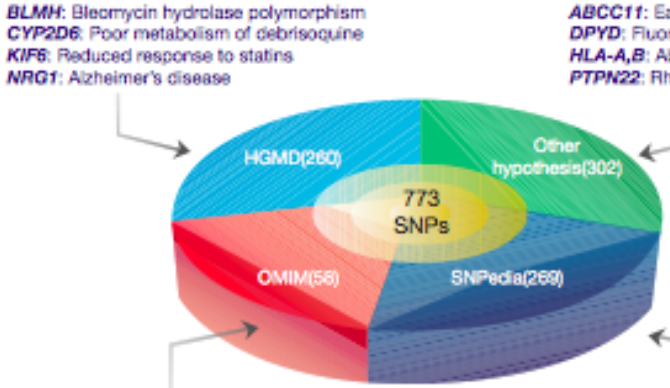


The Era of Personalized Genomes is Here!

A highly annotated whole-genome sequence of Korean individual

Nature, August 20, 2009, 460, 1011-1016





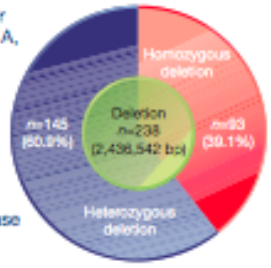
BLMH: Bleomycin hydrolase polymorphism
CYP2D6: Poor metabolism of debrisoquine
KIF6: Reduced response to statins
NRG1: Alzheimer's disease

ABCC11: Earwax related
DPYD: Fluorouracil toxicity
HLA-A,B: Abacavir hypersensitivity
PTPN22: Rheumatoid arthritis

ADD1: Salt-sensitive hypertension
ADRB2: Asthma, reduced β-2 agonist efficacy
LOXL1: Glaucoma
MMP9: Lumbar disc herniation
MTHFD1: Neural tube defects
NDUFV2: Parkinson's disease
SLC30A8: NIDDM
SP110: TB infection
UGT1A1: Gilbert's syndrome

Higher type II DM risk
Higher prostate cancer risk
Better rosiglitazone efficacy
Warfarin resistance
Better citalopram response

- ***ACOT1***: Acyl-CoA thioesterase 1
- ***ANGPT1***: Angiopoietin 1
- ***APBA1***: Amyloid β A4 precursor protein-binding, family A, member 1
- ***CARD6***: Caspase recruitment domain-containing protein 16
- ***FREM2***: Fras1-related extracellular matrix protein 2
- ***GCH1***: GTP cyclohydrolase 1
- ***GSTT2***: Glutathione S-transferase theta 2
- ***SGK1***: Serum/glucocorticoid-regulated kinase 1



- ***CAMTA1***: Calmodulin-binding transcription activator 1
- ***FRAS1***: Fraser syndrome 1
- ***LILRA3***: Leukocyte immunoglobulin-like receptor
- ***PADI4***: Peptidylarginine deiminase
- ***TP63***: Tumour protein p63
- ***TPH2***: Tryptophan hydroxylase 2
- ***TRY6***: Trypsinogen C

RESEARCH ARTICLE

PRENATAL DIAGNOSIS

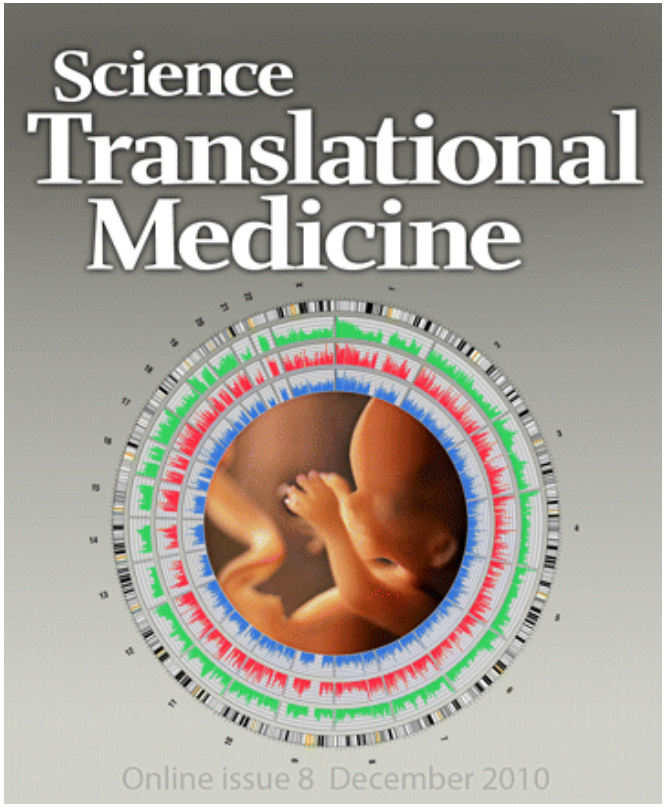
Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus

Science Translational Medicine, December 8, 2010 (61,1-12)

Sequencing DNA From the Blood of a Pregnant Woman Allows the Complete Genome Of the Fetus to Be Decoded!

~10% of DNA in Maternal Plasma is From the Fetus

A New Era in DNA Testing!!



February 11, 2010

Whole Genome of Ancient Human Is Decoded

Vol 463|11 February 2010|doi:10.1038/nature08835 nature

ARTICLES

Ancient human genome sequence of an extinct Palaeo-Eskimo

Morten Rasmussen[1,2]*, Yingrui Li[2,3]*, Stinus Lindgreen[1,4]*, Jakob Skou Pedersen[4], Anders Albrechtsen[4], Ida Moltke[4], Mait Metspalu[5], Ene Metspalu[5], Toomas Kivisild[5,6], Ramneek Gupta[7], Marcelo Bertalan[7], Kasper Nielsen[7], M. Thomas P. Gilbert[1,2], Yong Wang[8], Maanasa Raghavan[1,9], Paula F. Campos[1], Hanne Munkholm Kamp[1,4], Andrew S. Wilson[10], Andrew Gledhill[10], Silvana Tridico[11,12], Michael Bunce[12], Eline D. Lorenzen[1], Jonas Binladen[1], Xiaosen Guo[2,3], Jing Zhao[2,3], Xiuqing Zhang[2,3], Hao Zhang[2,3], Zhuo Li[2,3], Minfeng Chen[2,3], Ludovic Orlando[13], Karsten Kristiansen[2,3,4], Mads Bak[14], Niels Tommerup[14], Christian Bendixen[15], Tracey L. Pierre[16], Bjarne Grønnow[17], Morten Meldgaard[18], Claus Andreasen[19], Sardana A. Fedorova[5,20], Ludmila P. Osipova[21], Thomas F. G. Higham[9], Christopher Bronk Ramsey[10], Thomas v. O. Hansen[22], Finn C. Nielsen[22], Michael H. Crawford[23], Søren Brunak[7,24], Thomas Sicheritz-Pontén[7], Richard Villems[5], Rasmus Nielsen[4,8], Anders Krogh[2,4], Jun Wang[2,3,4] & Eske Willerslev[1,2]

We report here the genome sequence of an ancient human. Obtained from ~4,000-year-old permafrost-preserved hair, the genome represents a male individual from the first known culture to settle in Greenland. Sequenced to an average depth of 20×, we recover 79% of the diploid genome, an amount close to the practical limit of current sequencing technologies. We identify 353,151 high-confidence single-nucleotide polymorphisms (SNPs), of which 6.8% have not been reported previously. We estimate raw read contamination to be no higher than 0.8%. We use functional SNP assessment to assign possible phenotypic characteristics of the individual that belonged to a culture whose location has yielded only trace human remains. We compare the high-confidence SNPs to those of contemporary populations to find the populations most closely related to the individual. This provides evidence for a migration from Siberia into the New World some 5,500 years ago, independent of that giving rise to the modern Native Americans and Inuit.

From 5,000 Year-Old Hair!

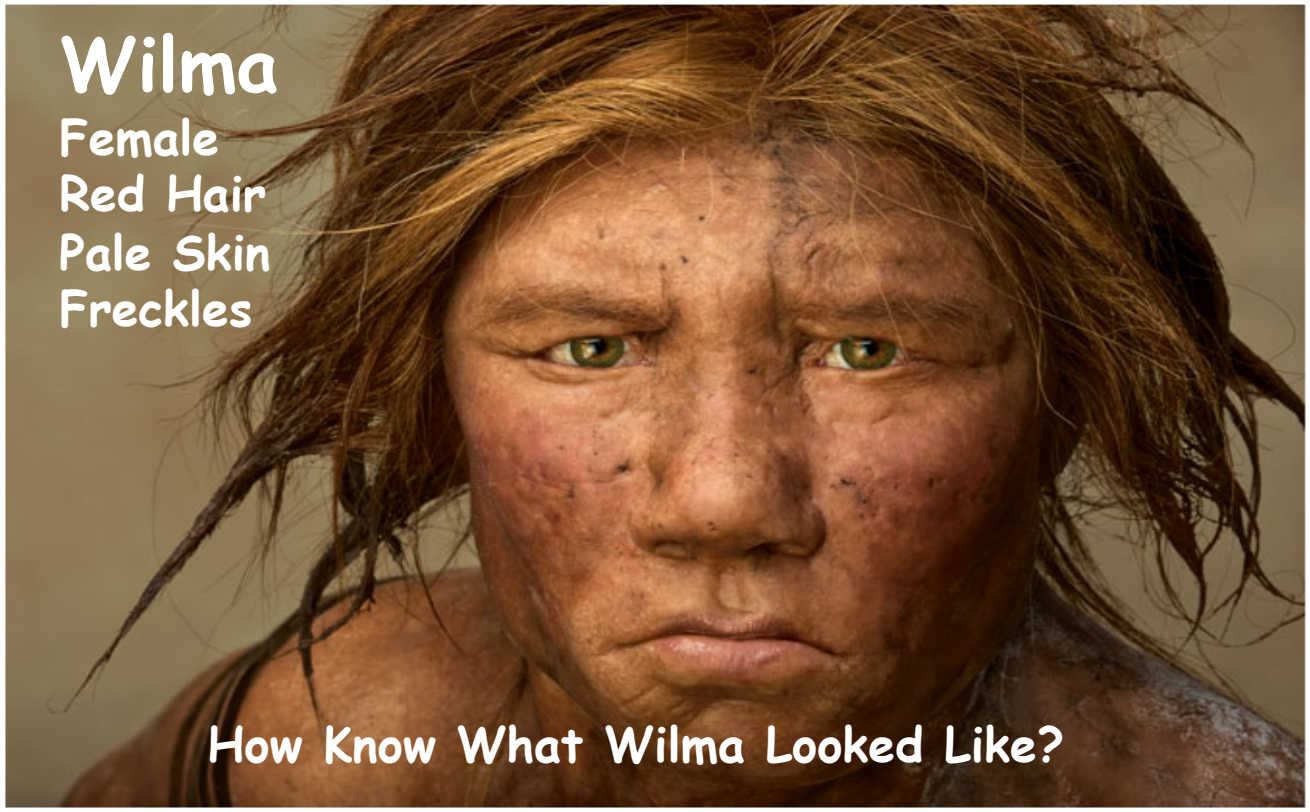
How Determine Phenotype?!

RESEARCH ARTICLE

Science, May 7, 2010 (328, 710-722)

A Draft Sequence of the Neandertal Genome

From a 45,000 Year-Old Bone



Reconstruction by Kennis & Kennis / Photograph by Joe McNally

For the first time, a Neanderthal female peers from the past in a reconstruction informed by both fossil anatomy and ancient DNA. At least some of her kind carried a gene for red hair and pale skin.

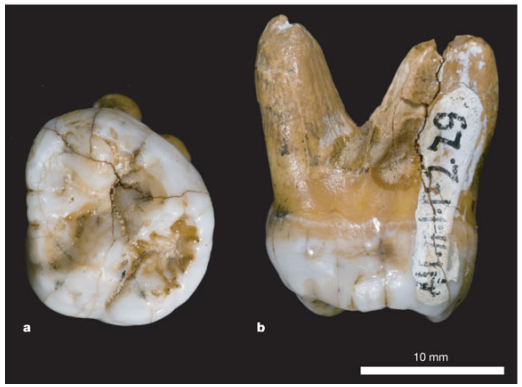
Nature, December 30, 2010 (468,1053-1060)

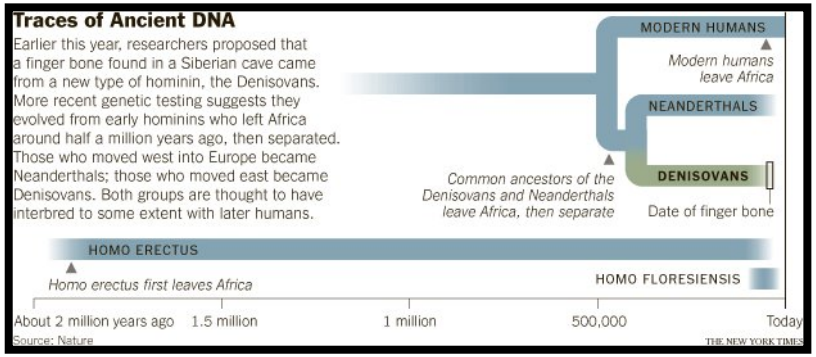
Genetic history of an archaic hominin group from Denisova Cave in Siberia

David Reich[1,2]*, Richard E. Green[3,4]*, Martin Kircher[3]*, Johannes Krause[3,5]*, Nick Patterson[2]*, Eric Y. Durand[6]*, Bence Viola[3,7]*, Adrian W. Briggs[1,3], Udo Stenzel[3], Philip L. F. Johnson[8], Tomislav Maricic[3], Jeffrey M. Good[9], Tomas Marques-Bonet[10,11], Can Alkan[10], Qiaomei Fu[3,12], Swapan Mallick[1,2], Heng Li[2], Matthias Meyer[3], Evan E. Eichler[10], Mark Stoneking[3], Michael Richards[7,13], Sahra Talamo[7], Michael V. Shunkov[14], Anatoli P. Derevianko[14], Jean-Jacques Hublin[7], Janet Kelso[3], Montgomery Slatkin[6] & Svante Pääbo[3]

Using DNA extracted from a finger bone found in Denisova Cave in southern Siberia, we have sequenced the genome of an archaic hominin to about 1.9-fold coverage. This individual is from a group that shares a common origin with Neanderthals. This population was not involved in the putative gene flow from Neanderthals into Eurasians; however, the data suggest that it contributed 4–6% of its genetic material to the genomes of present-day Melanesians. We designate this hominin population 'Denisovans' and suggest that it may have been widespread in Asia during the Late Pleistocene epoch. A tooth found in Denisova Cave carries a mitochondrial genome highly similar to that of the finger bone. This tooth shares no derived morphological features with Neanderthals or modern humans, further indicating that Denisovans have an evolutionary history distinct from Neanderthals and modern humans.

DNA Sequence From 40,000 Year Old Fossil DNA!!

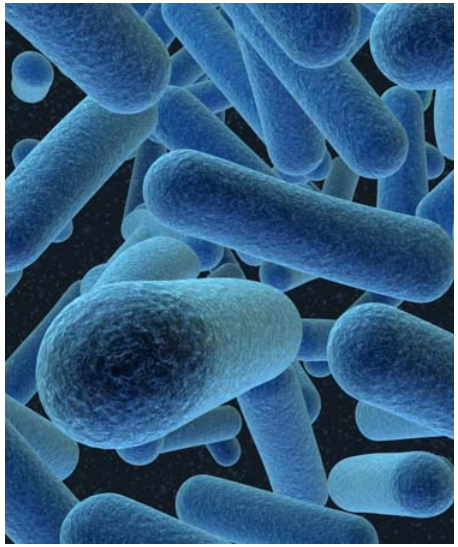


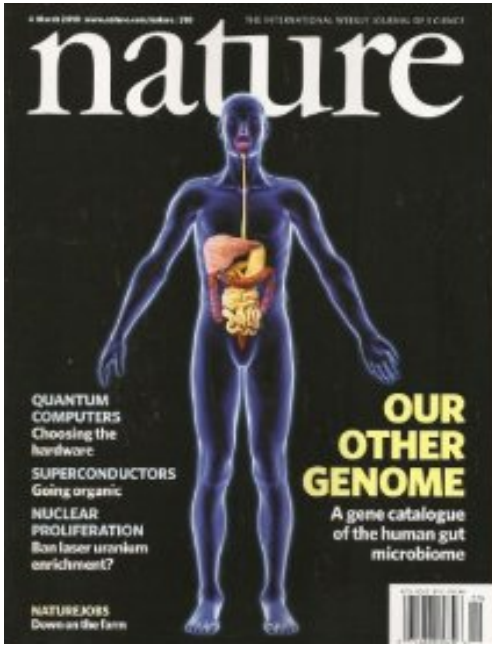


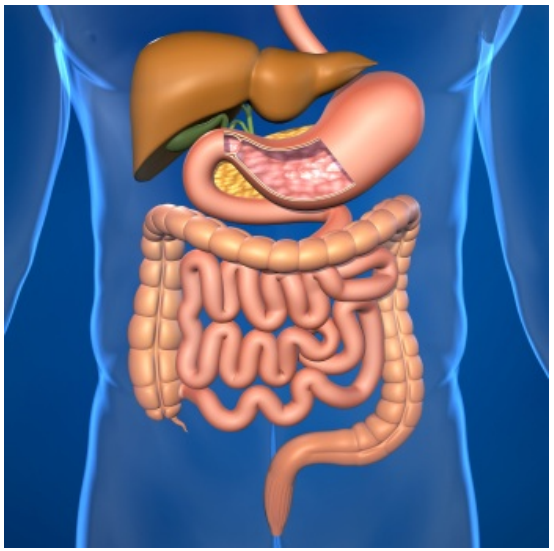
Nature, March 4, 2010, 464, 59-65

A human gut microbial gene catalogue established by metagenomic sequencing

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.







The 1,000 Genomes Project Will Provide Novel Insight in Human Genomes, Ancestry, & Disease Genes

Only Possible Using New Sequencing Methods

1,000 Genomes

Gene-sequencing projects keep getting bigger.
Tuesday, January 22, 2008
By Emily Singer

In a testament to the steady plummet in sequencing costs, today the National Human Genome Research Institute (NHGRI) announced a massive international collaboration to sequence the genomes of 1,000 people from around the world.

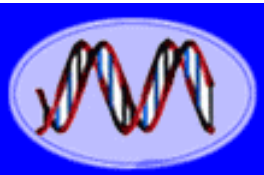
According to the NHGRI statement,

> "The 1000 Genomes Project will examine the human genome at a level of detail that no one has done before," said Richard Durbin, Ph.D., of the Wellcome Trust Sanger Institute, who is co-chair of the consortium. "Such a project would have been unthinkable only two years ago. Today, thanks to amazing strides in sequencing technology, bioinformatics and population genomics, it is now within our grasp. So we are moving forward to build a tool that will greatly expand and further accelerate efforts to find more of the genetic factors involved in human health and disease."
>
> During its two-year production phase, the 1000 Genomes Project will deliver sequence data at an average rate of about 8.2 billion bases per day, the equivalent of more than two human genomes every 24 hours. The volume of data--and the interpretation of those data--will pose a major challenge for leading experts in the fields of bioinformatics and statistical genetics.

2 Human Genomes Every 24 hrs !

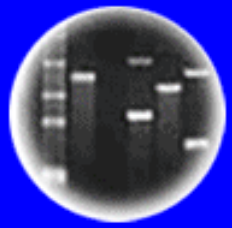
The 1,000 volunteers will be selected from those who participated in the HapMap project, a map of common genetic variation (see "A New Map for Health"), and will include:

Yoruba in Ibadan, Nigeria; Japanese in Tokyo; Chinese in Beijing; Utah residents with ancestry from northern and western Europe; Luhya in Webuye, Kenya; Maasai in Kinyawa, Kenya; Toscani in Italy; Gujarati Indians in Houston; Chinese in metropolitan Denver; people of Mexican ancestry in Los Angeles; and people of African ancestry in the southwestern United States.



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

ARTICLE Nature, October 10, 2010

doi:10.1038/nature09534

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of *de novo* germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

- Sequenced Genomes of ~900 individuals
- From Seven Different Global Populations
- Found 250-300 Loss-Of-Function Mutations (KOs) Per Person
- 10^{-8} bp Mutations per Generation (30 per Genome)
- 15,000,000 SNPs

Genome-Wide Testing Can Lead to "Surprises"

Identification of incestuous parental relationships by SNP-based DNA microarrays

The Lancet, February 12, 2011
377, 555-556

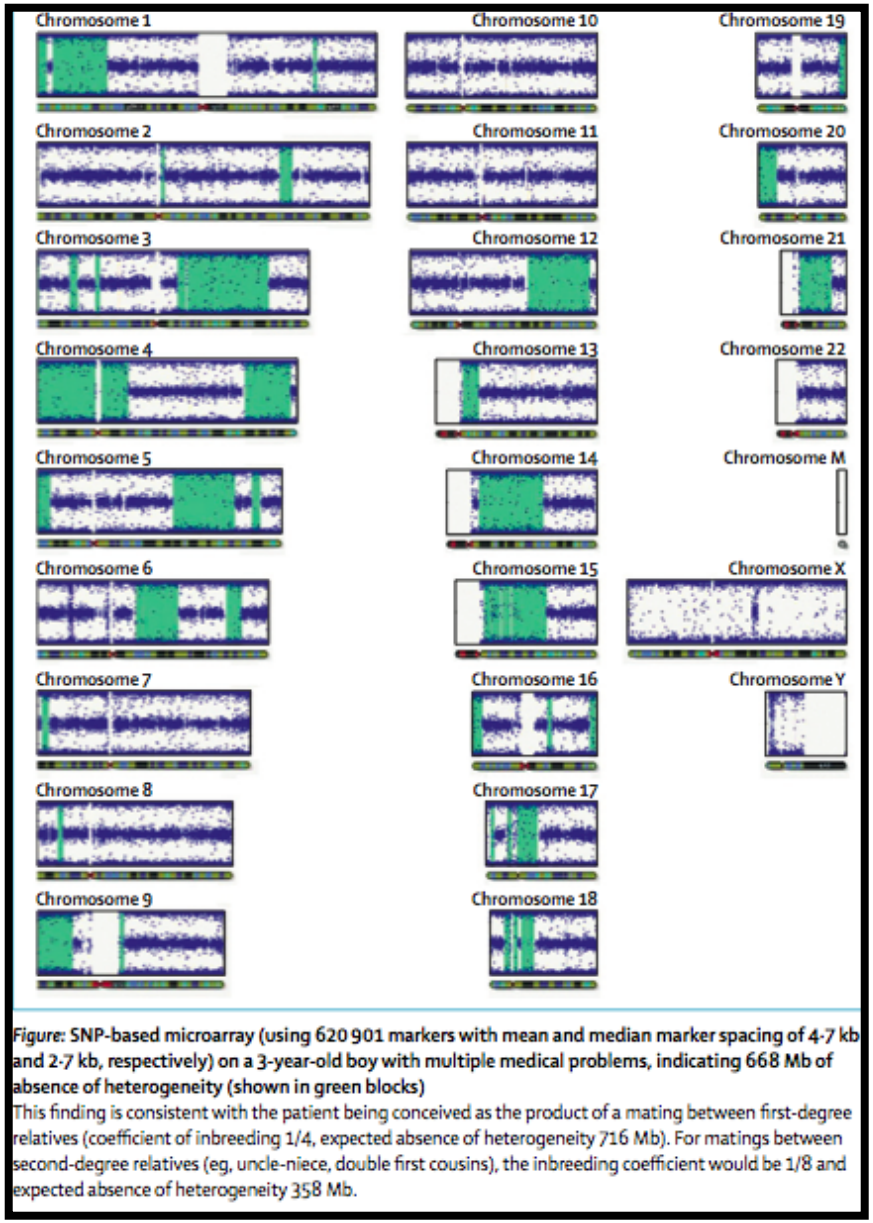
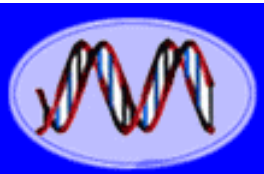
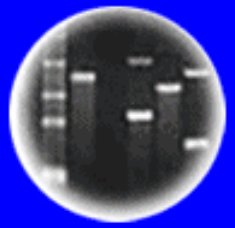


Figure: **SNP-based microarray (using 620 901 markers with mean and median marker spacing of 4·7 kb and 2·7 kb, respectively) on a 3-year-old boy with multiple medical problems, indicating 668 Mb of absence of heterogeneity (shown in green blocks)**
This finding is consistent with the patient being conceived as the product of a mating between first-degree relatives (coefficient of inbreeding 1/4, expected absence of heterogeneity 716 Mb). For matings between second-degree relatives (eg, uncle-niece, double first cousins), the inbreeding coefficient would be 1/8 and expected absence of heterogeneity 358 Mb.



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



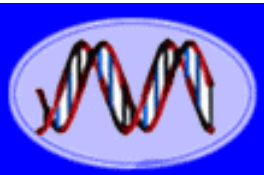
DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

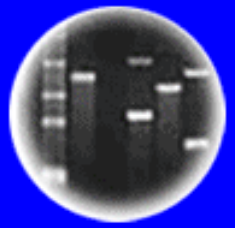
Genetic Screening Issues

- Why Screen For Genes?
- When is a Test Accurate Enough?
- Mandatory or Voluntary Screening?
- Who Should Be Tested?
- Employer & Insurance Company Testing?
- Protection From Genotype Discrimination?
- Testing for Genetic Diseases With No Cures?
- How Ensure Privacy & Confidentiality?
- Obligations to Inform Others (Sibling) of Genetic Disorder?
- How Ensure Privacy & Confidentiality?
- Genetic Databases?



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

Laws That Affect DNA Testing

- Constitution-Article I Section 8.8
  Promote the General Welfare
- 4th Amendment-Searches & Seizures
- Amendment X-Powers Reserved to States
- Federal Criminal Statutes
- State Constitutions
- State Tort & Criminal Statutes

National Genetic Privacy Law

What is GINA?

The Genetic Information Nondiscrimination Act, or GINA, is U.S. federal legislation with bipartisan support that protects Americans from discrimination (in health insurance and employment decisions) on the basis of genetic information. GINA has passed through Congress and was signed into law by the President on May 21, 2008. As a result, American insurance companies and health plans (including both group and individual insurers, as well as federally-regulated plans) will be prohibited from:

- looking at your predictive genetic information or genetic services before you enroll;
- "requesting or requiring" that you or your family members take a genetic test;
- restricting enrollment based on genetic information;
- changing your premiums based on genetic information.

GINA also prohibits U.S. employers (including employment agencies, labor organizations, and training programs) from:

- discriminating against who they hire or how much they pay on the basis of genetic information;
- "requesting or requiring" that you or your family members take a genetic test;
- disclosing your genetic information in their possession except under specific and specially controlled circumstances.

GINA does not cover life or disability insurance providers.

NCSL
National Conference of State Legislatures

California Genetic Laws

- Newborn Genetic Screening
- Genetic Non Discrimination in Insurance
- Human Cloning Laws
- Genetic Employment Laws
- Genetic Counselor Licensing Laws
- Embryonic and Fetal Research Laws
- Embryo and Gamete Disposition Laws

NCSL
NATIONAL CONFERENCE of STATE LEGISLATURES

California Genetic Privacy Laws

State and Statute	Personal Access to Genetic Information Required	Consent Required to				Define as Personal Property		Specific Penalties for Genetic Privacy Violations
		Perform/ Require Genetic Test	Obtain/ Access Genetic Information	Retain Genetic Information	Disclose Genetic Information	Genetic Information	DNA Samples	
Alabama								
Alaska §18.13.010-100		x	x	x	x	x	x	x
Arizona §20-448.02		x			x			
Arkansas §20-35-101 to 103					x			
California Insurance §10149.1					x			x

California Department of Public Health

CDPH

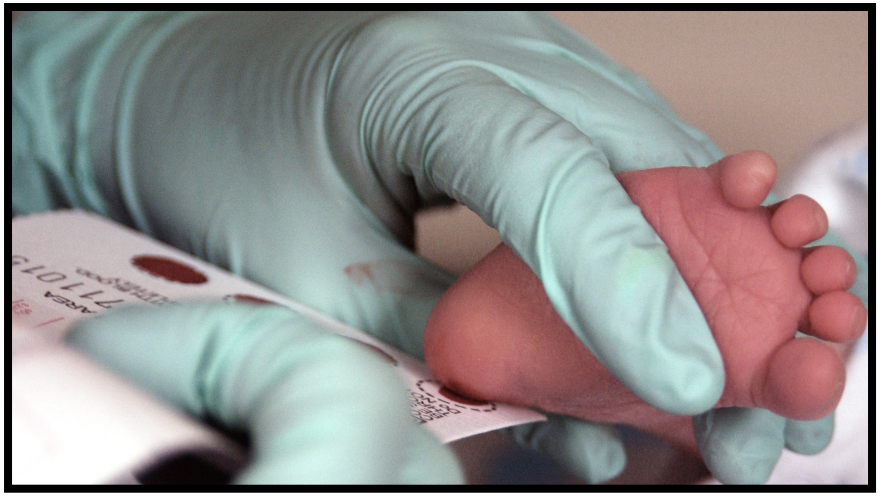
Cystic Fibrosis
Metabolic Disorders
Blood Disorders

~80 Total

Newborn-Blood Storage Law Stirs Fears of DNA Warehouse

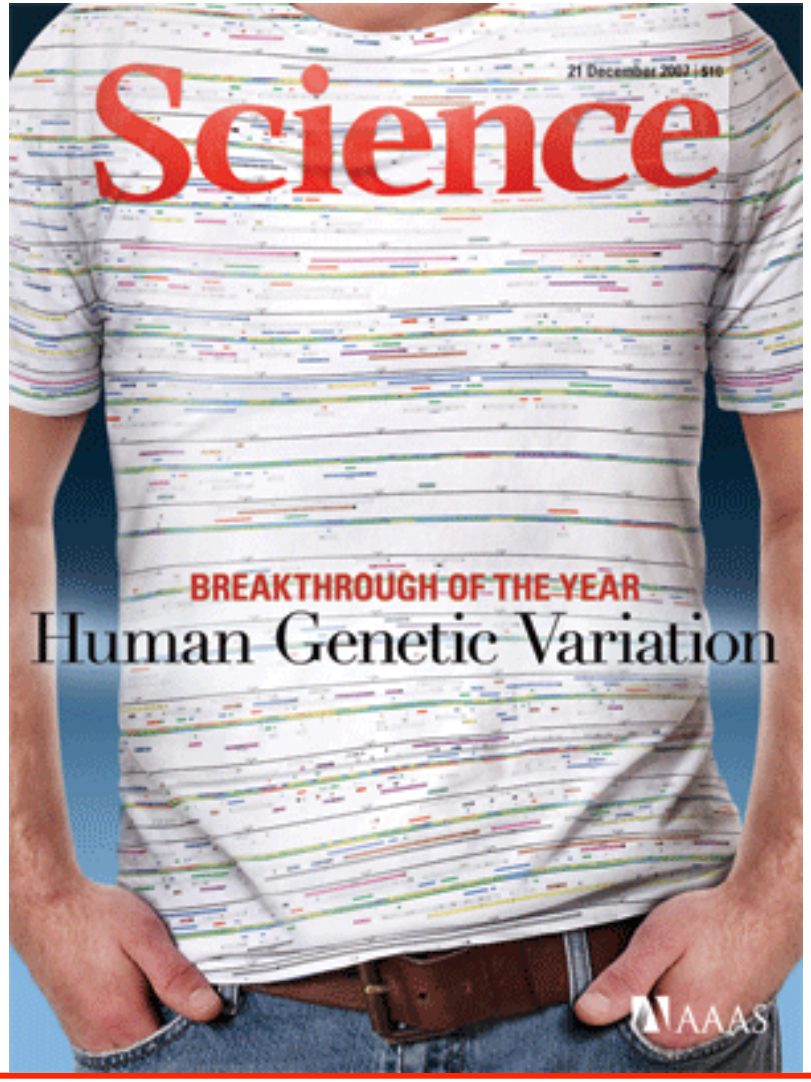
By Alexis Madrigal 05.21.08

An obscure bill that sailed through Congress and was signed into law last month is stoking fears of a nationwide DNA warehouse potentially open to abuse by law enforcement agencies or health insurance companies.



There Are Large DNA Sequence Variations in Human Populations

Variation in Genes (e.g., Disease Genes) Accounts For Only a Small Amount of Human DNA Variation



DNA Sequence Variation Makes us Individuals! Genetic Variability-Allelic Differences

Identifying SNPs in the Human Genome

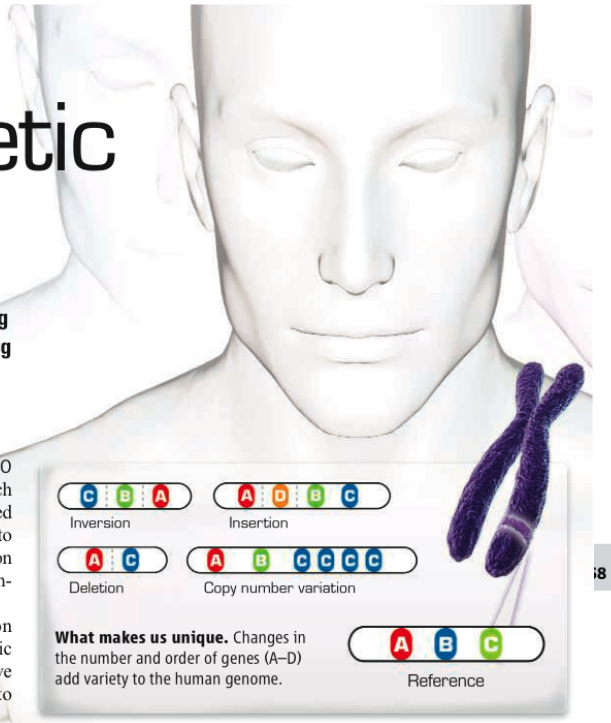
BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.



What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

Remember: Most SNPs Are Not in Gene Coding Regions

nature

ARTICLES

Identify From Sequencing the Genome Regions (and soon Genomes) of Individuals From Different Groups

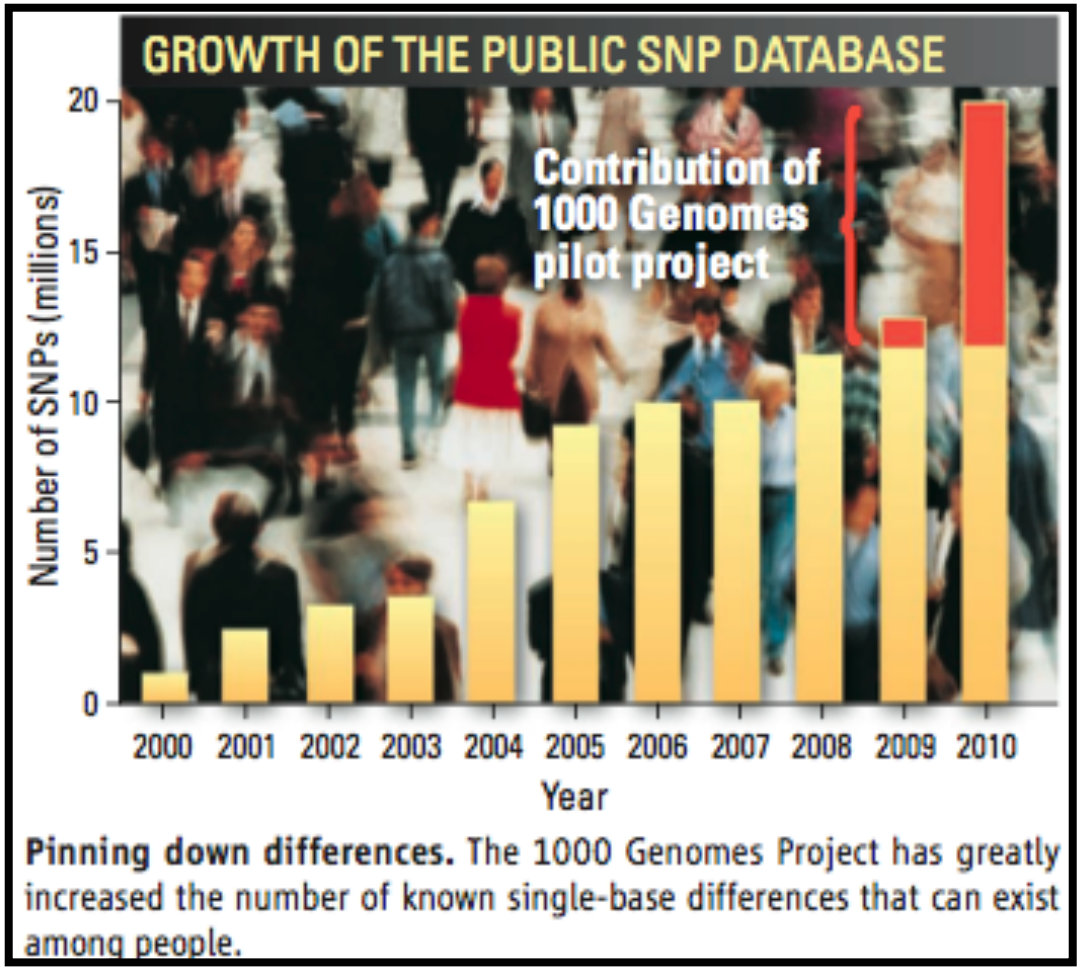
A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination rates vary systematically around genes and between genes of different function. Finally, we demonstrate increased differentiation at non-synonymous, compared to synonymous, SNPs, resulting from systematic differences in the strength or efficacy of natural selection between populations.

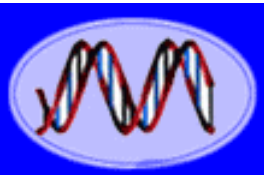
GENOMICS

1000 Genomes Project Gives New Map Of Genetic Diversity



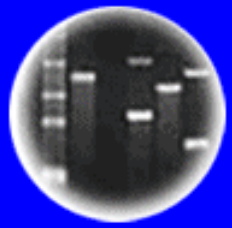
Pinning down differences. The 1000 Genomes Project has greatly increased the number of known single-base differences that can exist among people.

20,000,000 SNPs and Growing!



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

ARTICLE Nature, October 10, 2010

doi:10.1038/nature09534

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of *de novo* germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

- Sequenced Genomes of ~900 individuals
- From Seven Different Global Populations
- Found 250-300 Loss-Of-Function Mutations (KOs) Per Person
- 10^{-8} bp Mutations per Generation (30 per Genome)
- 15,000,000 SNPs

Alleles And Homologous Chromosomes-A Reminder

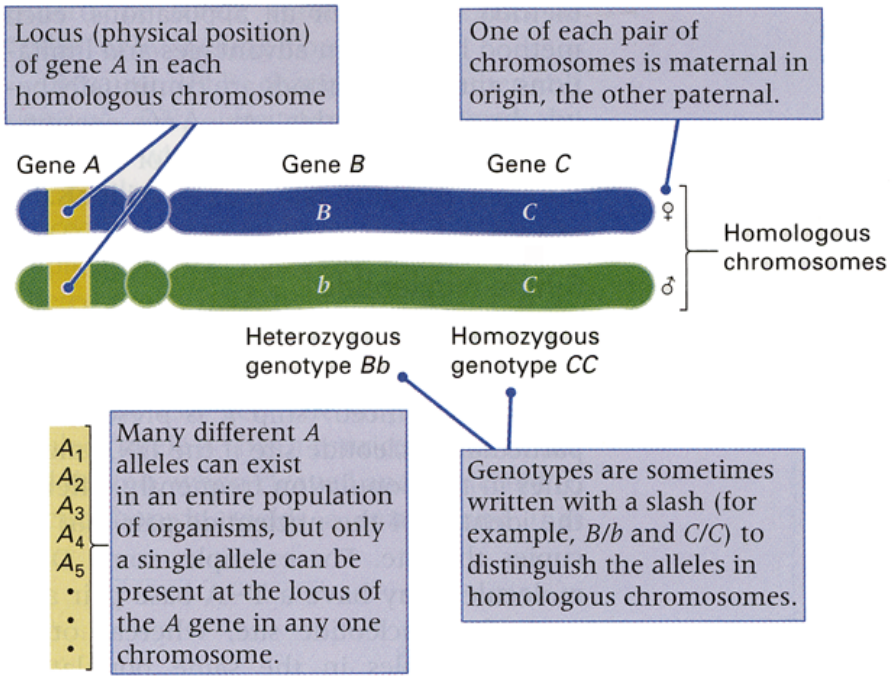


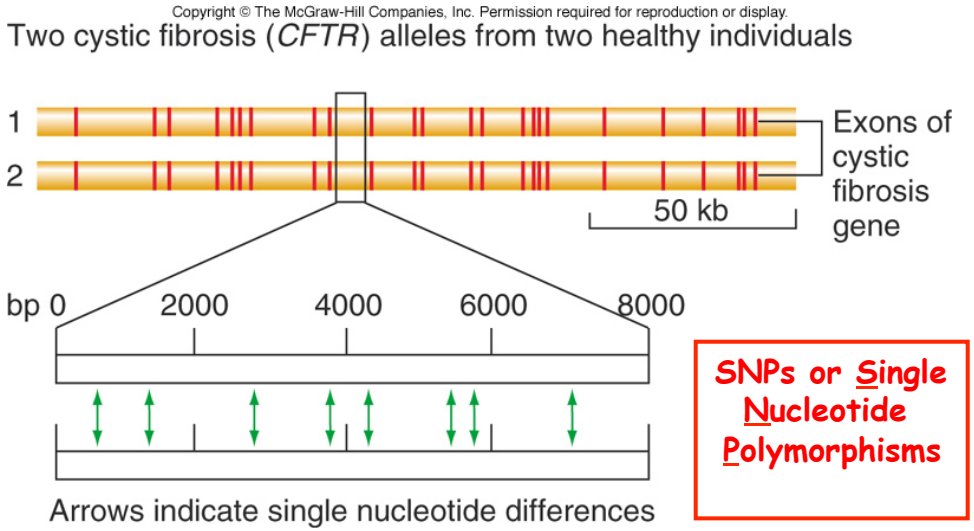
Figure 2.22 Key concepts and terms used in modern genetics. Note that a single gene can have any number of alleles in the population as a whole, but no more than two alleles can be present in any one individual.

Marker

Individuals May Contain Two Different Alleles at any DNA Location

There can be an Infinite # of Alleles for any Gene (or DNA sequence in a Population

Most DNA Variations Between Individuals Occur Because of Base-Pair Changes in Non-Coding Regions of the Genome



SNPs or Single Nucleotide Polymorphisms

To Be on the safe side, suppose you assume that only 80% (0.8) of the 3 billion base pairs in the genome are noncoding, and on average only 1 base pair in 700 is polymorphic. With these assumptions, you can determine the frequency of polymorphism within a single individual by multiplying 3 billion by 0.8 and then multiplying that amount by 1/700:

(3x109) x 0.8 = 2.4 x 109, (2.4 x 109) x 1/700 = 3.4 million.

The result of 3.4 million is astonishing: It means that there are millions of differences between any two haploid sets of human chromosomes. Combined with differences in coding and regulatory sequences (which occur much less frequently), the millions of polymorphisms at anonymous loci contribute to an enormous pool of potential DNA markers.

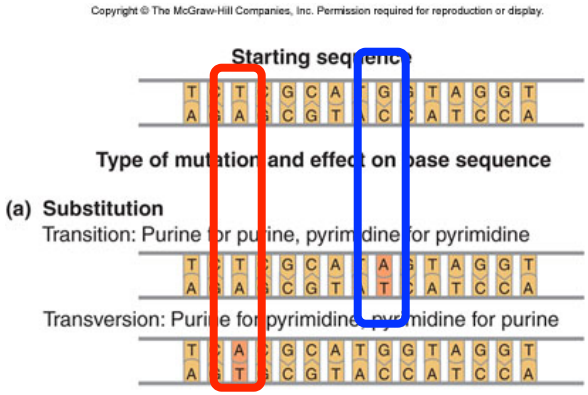
TABLE 11.1 Classes of DNA Polymorphisms

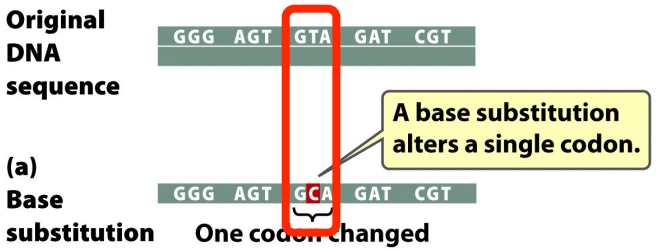
Class	Size of Locus	Number of Alleles	Number of Loci in Population	Rate of Mutation	Use	Method of Detection
SNP	Single base pair	2	100 million	10^{-9}	Linkage and association mapping	PCR followed by ASO hybridization or primer extension
Microsatellite	30–300 bp	2–10	200,000	10^{-3}	Linkage and association mapping	PCR and gel electrophoresis
Multilocus minisatellite	1–20 kb	2–10	30,000	10^{-3}	DNA fingerprinting	Southern blot and hybridization
Indels (deletions and duplications)	1–100 bp	2	N/A	$<10^{-9}$	Linkage and association mapping	PCR and gel electrophoresis

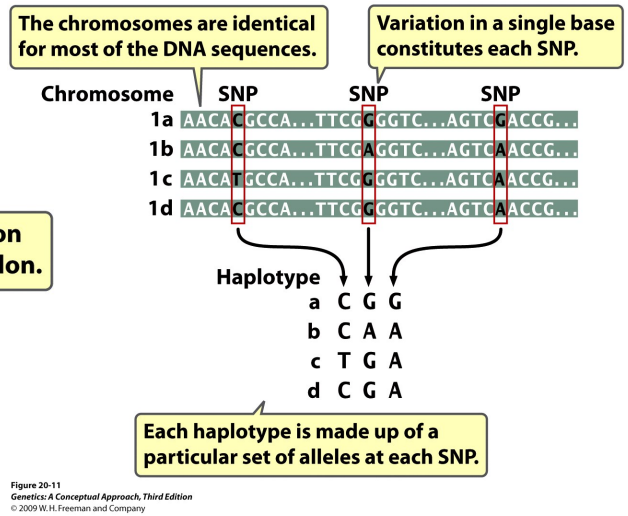
This is What Makes Us Unique Individuals!

There is ~1bp Change per 700bp in Human Genomes or ~3.4 Million bp Differences Between Individuals ~0.1% of Genome

How Do SNPs Arise in the Human Genome During DNA Replication?







Most SNPs are Single Nucleotide Changes that Have No Effect on the Phenotype or Gene Function! They Are Outside Coding Sequence of Genes -- Between Genes or in Introns

Different "Forms" of the Same SNP = Allele!

DNA Sequence Changes in the Genome Are Rare

TABLE 9.1 Five Classes of DNA Polymorphism

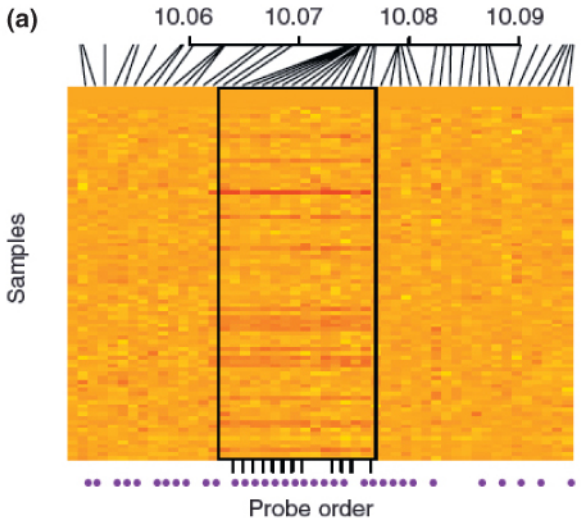
Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
Single base (SNP)	Mutagens or replication errors	10^{-8}–10^{-9}	1/700 bp	3 million
Microsatellite (VNTR or SSR)	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers	10^{-3}	Unknown; discovered by chance	Fewer than 100 families known, yielding 1000 copies in all
Deletions	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Duplications	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Other insertions (excluding those resulting from micro- or minisatellite recombination)	Transposable elements	Extremely rare	Very low	0 – a few
Complex haplotype (any locus of 5 kb or more)	Any of the above	Combination of the above	Not applicable	Not applicable

Only a Few Affect Gene Function & Lead to a Visible Mutation!

Categories of Human DNA Variants From Genome Sequencing

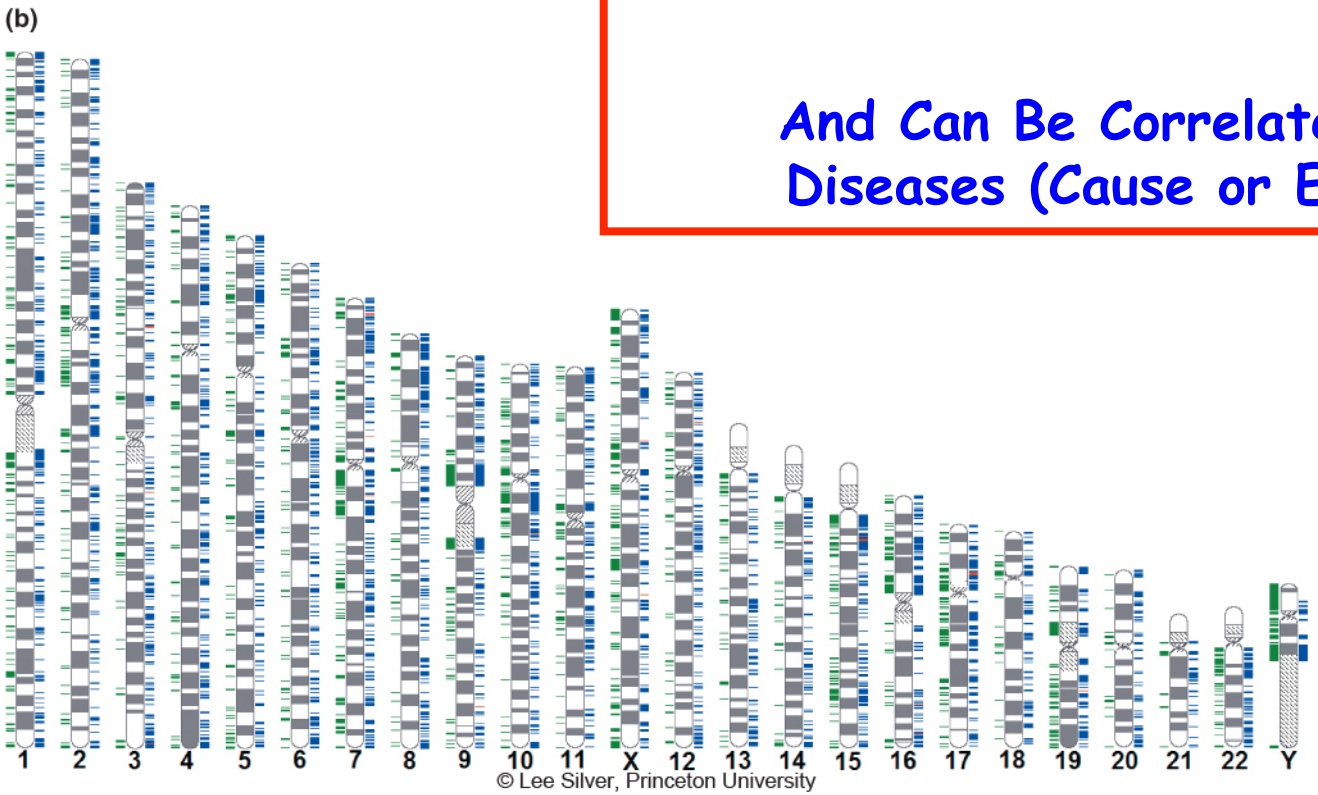
TABLE 11.1 Categories of Genetic Variants

	Short Name	Size	Frequency	Total Loci Recorded	Method of Detection			
					DNA Microarray	PCR & Gel Electrophoresis	PCR & ASO Hybridization	DNA Sequence
Single nucleotide polymorphism	SNP	1 bp	1 kb	18,000,000	Yes		Yes	
Insertion/deletion	InDel	2–100 bp	10 kb	200,000	(Yes)	Yes		
Simple sequence repeat	SSR, microsatellite	3–200 bp	30 kb	100,000		Yes		
Copy number variation/copy number polymorphism	CNV/CNP	0.1–1,000 kb	3 Mb	8,600	Yes			
Complex variant			500 kb					Yes

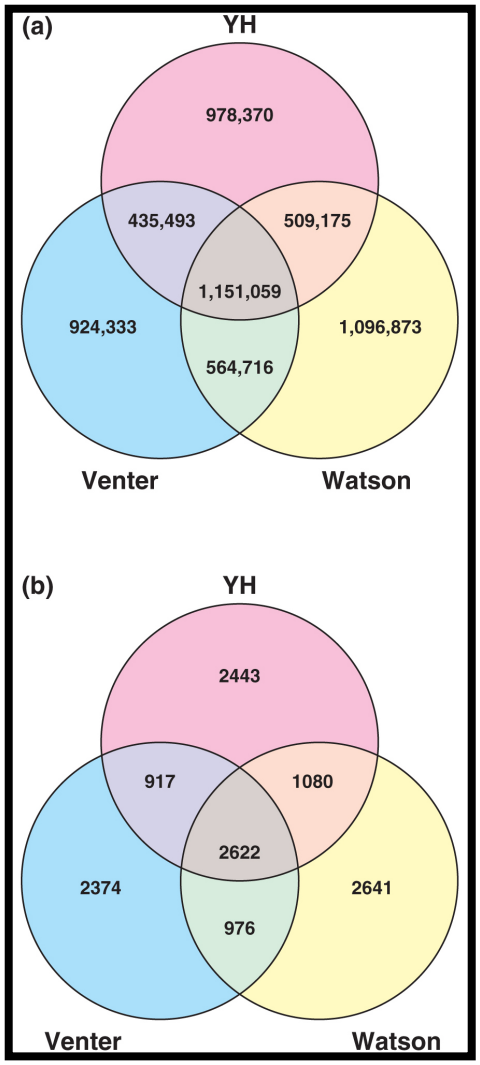


Copy Number Variants Also Occur in the Human Genome and Can Vary From Individual to Individual

And Can Be Correlated With Diseases (Cause or Effect?)



Comparison of SNPs in Watson's and Ventner's Genomes



YH= Anonymous Chinese Man

(a)

Allele	Genome	Single-Strand Notation
1	AGCTCTGATAC TCGAGACTATG	AGCTCTGATAC
2	AGCTCCGATAC TCGAGGCTATG	AGCTCCGATAC

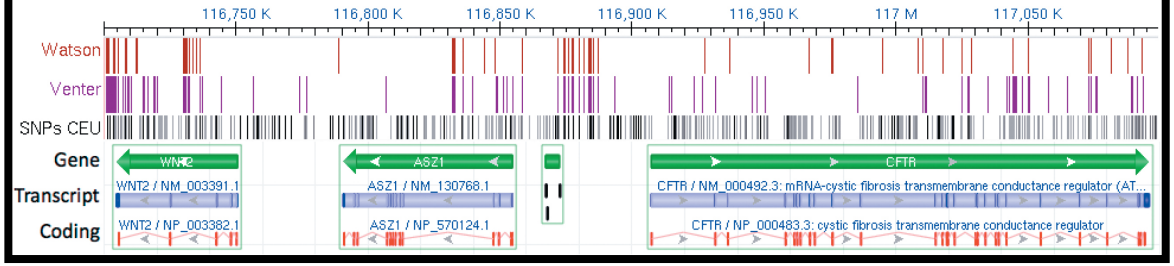
(b)

Genotype	Notation
Homozygous T	TT
Heterozygous T	TC
Homozygous C	CC

(c)

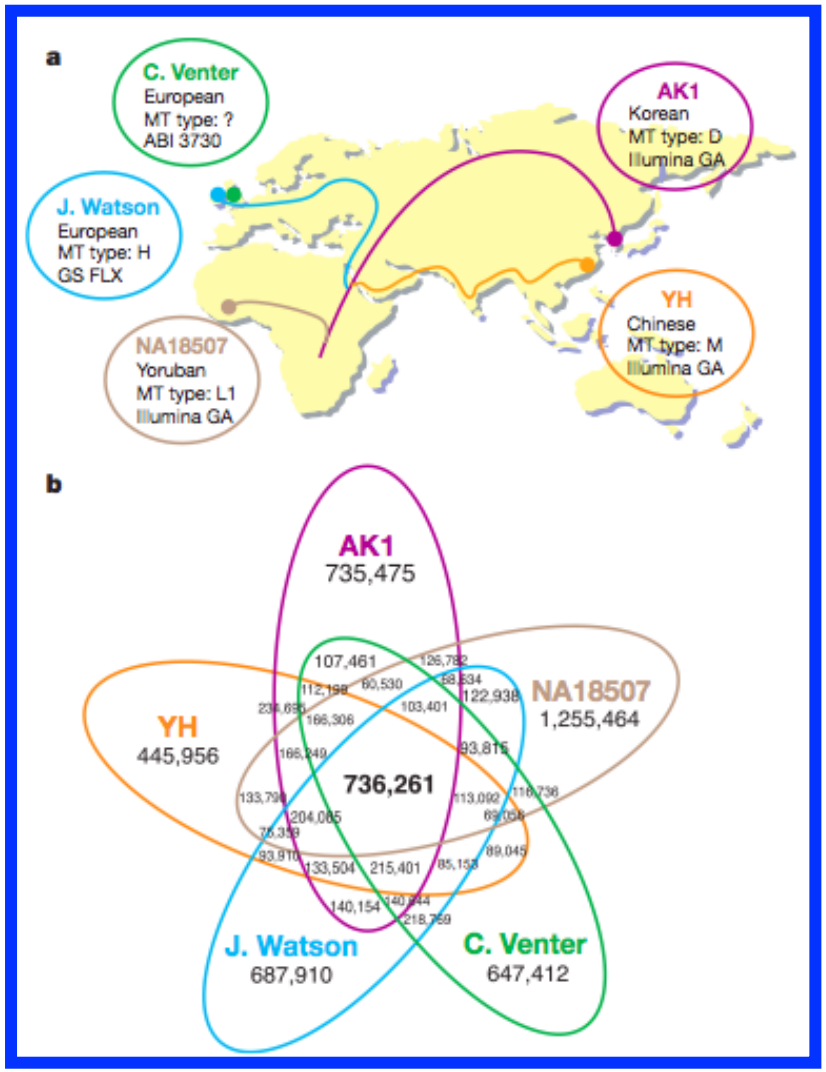
Human 1 TTGACGTATAAATGATCTTTATATTTTCAGAAGTC
Human 2 TTGACGTATAAATGATCTTTATATCTTCAGAAGTC
Chimp TTGACATATAAATGATCTTTATATCTTCAGAAGTC

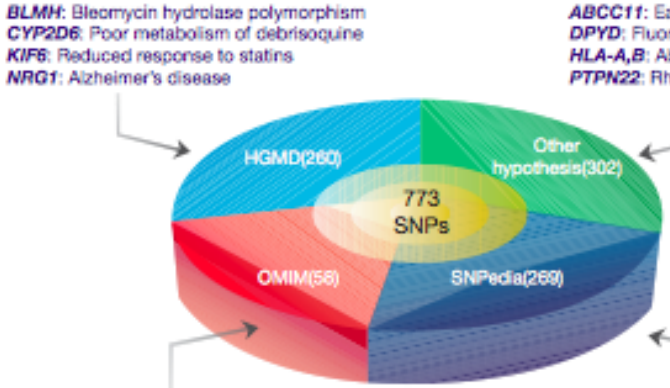
(d)



A highly annotated whole-genome sequence of Korean individual

Nature, August 20, 2009, 460, 1011-1016





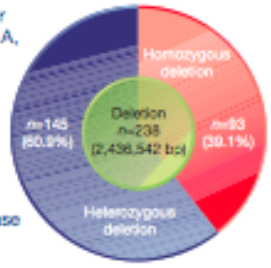
BLMH: Bleomycin hydrolase polymorphism
CYP2D6: Poor metabolism of debrisoquine
KIF6: Reduced response to statins
NRG1: Alzheimer's disease

ABCC11: Earwax related
DPYD: Fluorouracil toxicity
HLA-A,B: Abacavir hypersensitivity
PTPN22: Rheumatoid arthritis

ADD1: Salt-sensitive hypertension
ADRB2: Asthma, reduced β-2 agonist efficacy
LOXL1: Glaucoma
MMP9: Lumbar disc herniation
MTHFD1: Neural tube defects
NDUFV2: Parkinson's disease
SLC30A8: NIDDM
SP110: TB infection
UGT1A1: Gilbert's syndrome

Higher type II DM risk
Higher prostate cancer risk
Better rosiglitazone efficacy
Warfarin resistance
Better citalopram response

- ***ACOT1***: Acyl-CoA thioesterase 1
- ***ANGPT1***: Angiopoietin 1
- ***APBA1***: Amyloid β A4 precursor protein-binding, family A, member 1
- ***CARD6***: Caspase recruitment domain-containing protein 16
- ***FREM2***: Fras1-related extracellular matrix protein 2
- ***GCH1***: GTP cyclohydrolase 1
- ***GSTT2***: Glutathione S-transferase theta 2
- ***SGK1***: Serum/glucocorticoid-regulated kinase 1



- ***CAMTA1***: Calmodulin-binding transcription activator 1
- ***FRAS1***: Fraser syndrome 1
- ***LILRA3***: Leukocyte immunoglobulin-like receptor
- ***PADI4***: Peptidylarginine deiminase
- ***TP63***: Tumour protein p63
- ***TPH2***: Tryptophan hydroxylase 2
- ***TRY6***: Trypsinogen C

Using SNPs or DNA Sequence Variation As Markers For Disease Genes

Remember: Only a Small Fraction of Human Genes Are Known To Cause Diseases

DNA Tests Available For Many Known Disease Genes

Table 11.1 GENETIC DISEASE TESTING

Genetic Disease Condition	Genetic Basis for Disease and Symptoms
Cancers (brain tumors; urinary bladder, prostate, ovarian, breast, brain, lung, and colorectal cancers)	A variety of different mutant genes can serve as markers for genetic testing.
Cystic fibrosis	Large number of mutations in the CFTR (cystic fibrosis transmembrane conductance regulator) gene on chromosome 7. Causes lung infections and problems with pancreatic, digestive, and pulmonary functions.
Duchenne muscular dystrophy	Defective gene (dystrophin) on the X chromosome causes muscle weakness and muscle degeneration.
Familial hypercholesterolemia	Mutant gene on chromosome 19 causes extremely high levels of blood cholesterol.
Hemophilia	Defective gene on the X chromosome makes it difficult for blood to clot when bleeding.
Huntington disease	Mutation in gene on chromosome 4 causes neurodegenerative disease in adults.
Phenylketonuria (PKU)	Mutation in gene required for converting the amino acid phenylalanine into the amino acid tyrosine. Causes severe neurological damage, including mental retardation.
Severe combined immunodeficiency (SCID)	Immune system disorder caused by mutation of the adenosine deaminase gene.
Sickle-cell disease	Mutation in β-globin gene on chromosome 11 affects hemoglobin structure and shape of red blood cells, which disrupts oxygen transport in blood and causes joint pain.
Tay-Sachs disease	Rare mutation of a gene on chromosome 5 causes certain types of lipids to accumulate in the brain. Causes paralysis, blindness, retardation, and respiratory infections.

RFLPs or DNA Markers (SNPs) Can Be Used to Follow/ Identify Gene Alleles if Linked

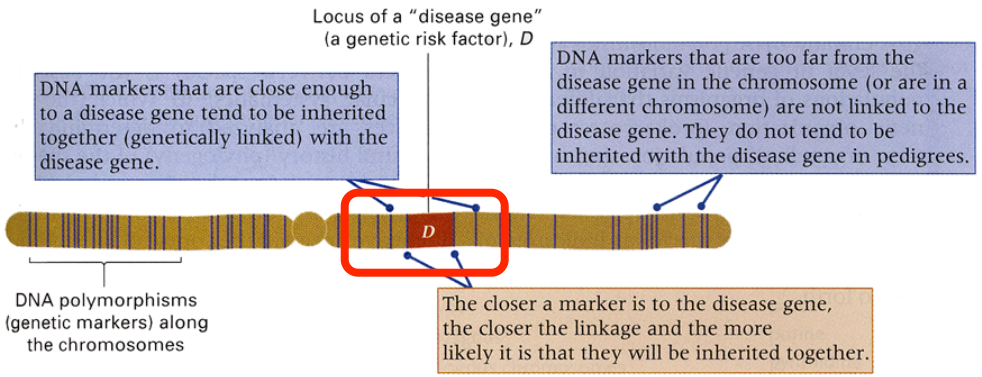


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is determined by examining the known genomic locations of the DNA polymorphisms that are linked with it.

Useful for DNA Testing & Genetic Diagnosis!

Markers For Wrinkled Peas

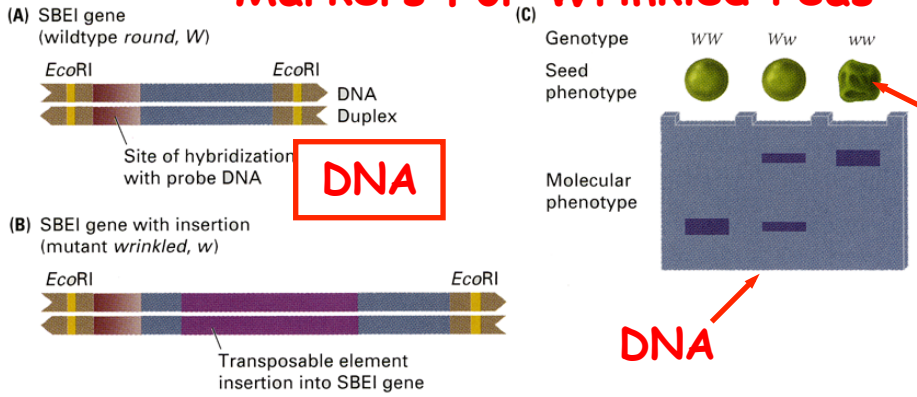
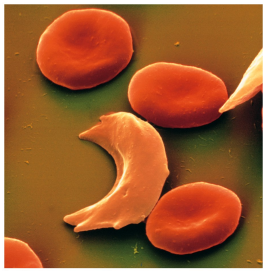
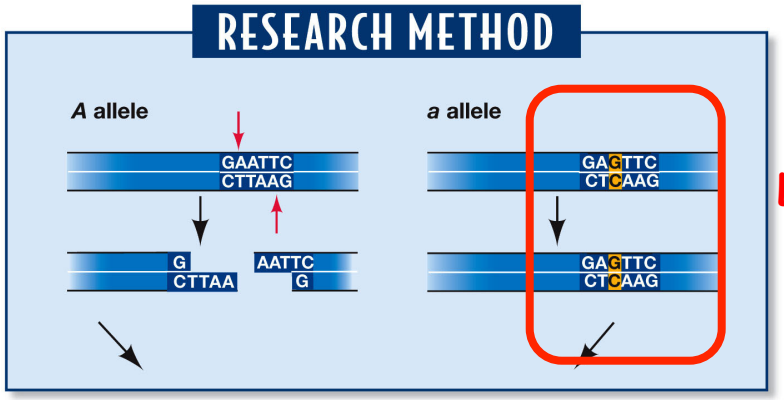


Figure 3.2 (A) *W* (round) is an allele of a gene that specifies the amino acid sequence of starch branching enzyme I (SBEI). (B) *w* (wrinkled) is an allele that encodes an inactive form of the enzyme because its DNA sequence is interrupted by the insertion of a transposable element. (C) At the level of the morphological phenotype, *W* is dominant to *w*: Genotype *WW* and *Ww* have round seeds, whereas genotype *ww* has wrinkled seeds. The molecular difference between the alleles can be detected as a restriction fragment length polymorphism (RFLP) using the enzyme *Eco*RI and a probe that hybridizes at the site shown. At the molecular level, the alleles are codominant: DNA from each genotype yields a different molecular phenotype—a single band differing in size for homozygous *WW* and *ww*, and both bands for heterozygous *Ww*.

Also Useful in Breeding and Obtaining Markers For Specific Traits!

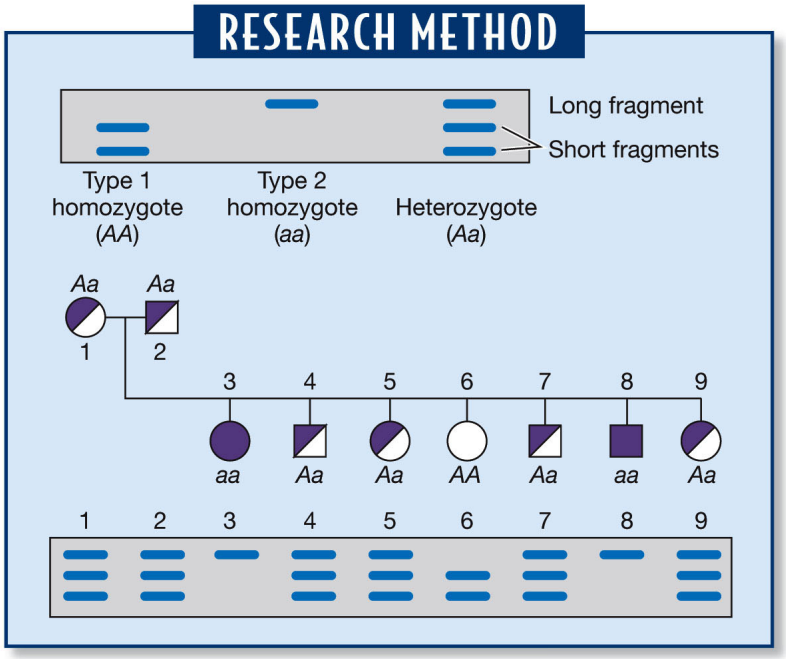


Using RFLPs + Markers to Identify the Sickle Cell Allele (Single Gene Test)



Loss of Restriction Site in a Allele (in gene)

Detected By Blots Or PCR



SNP Leads to RFLP!!!

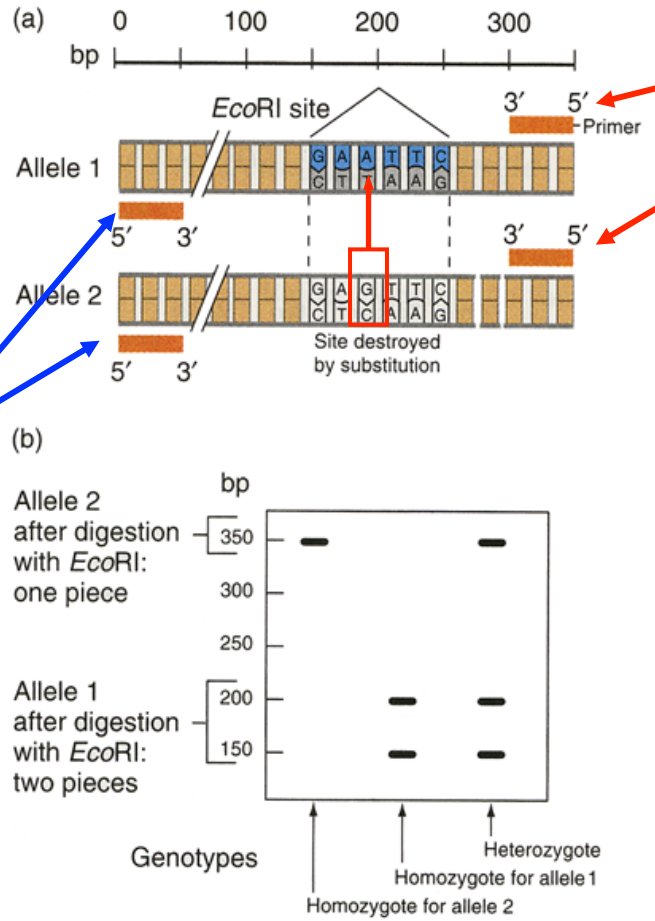
LIFE 8e, Figure 17.8 (Part 2)

Recall: PCR Can Be Used to Identify SNP-Generated RFLPs and DNA Variation

Identify by PCR Using Primers For Surrounding Conserved Sequences

Conserved Sequence 1

Conserved Sequence 2



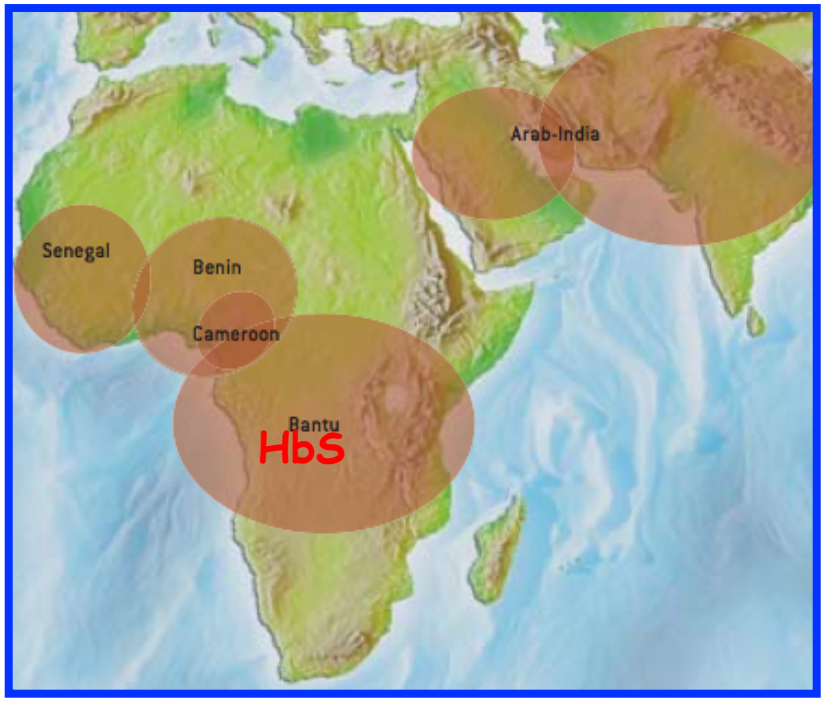
1. PCR
2. Restriction Digest
3. Gel

Figure 9.7 Restriction site polymorphisms can be detected most efficiently with PCR-based protocols. (a) PCR amplification

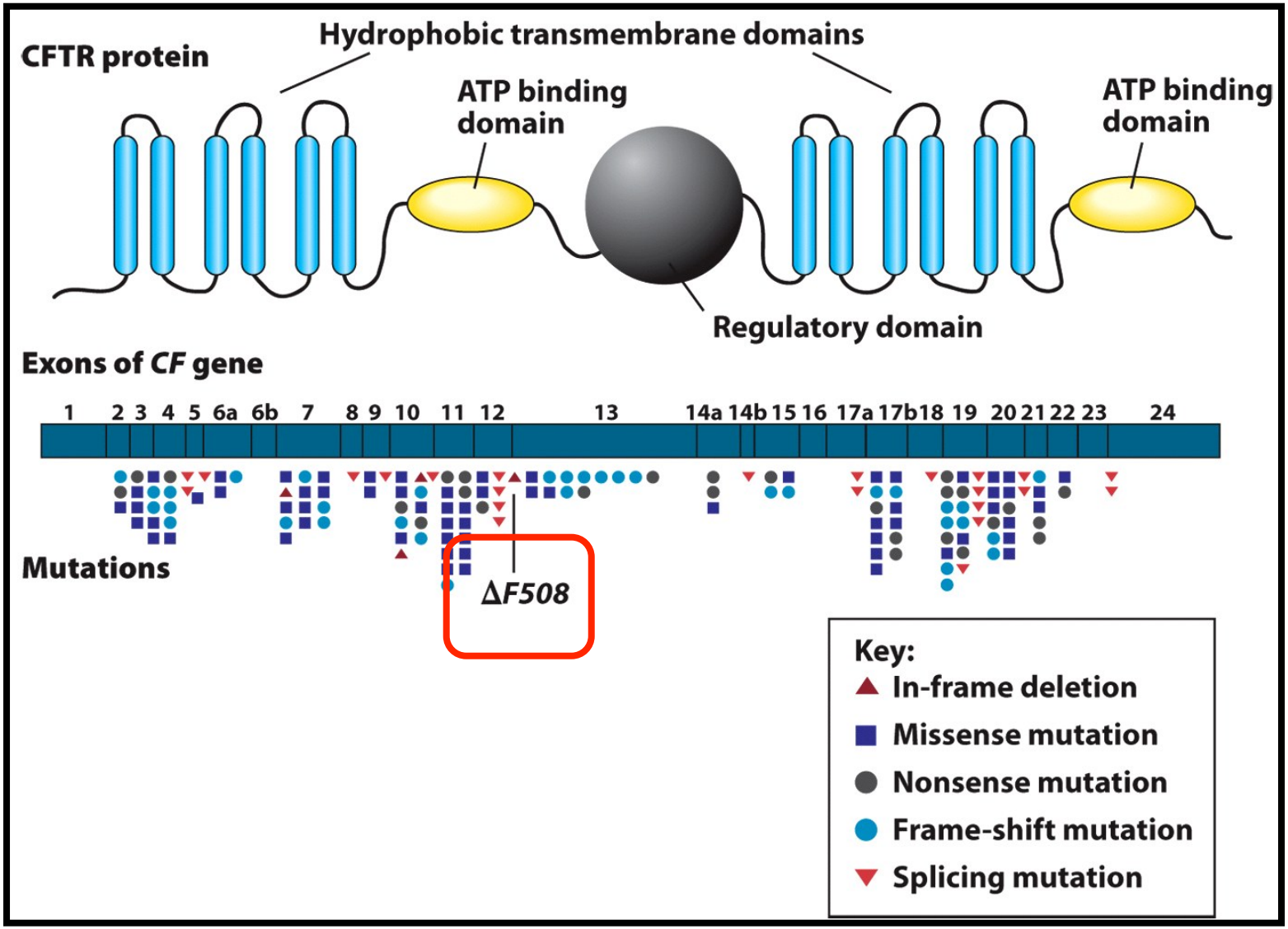
Sickle Cell Anemia & Cystic Fibrosis Are Examples of Founder Mutations

Noteworthy Founder Mutations

Affected gene	Condition	Mutation origin	Migration	Possible advantage of one copy
HFE	Iron overload	Far northwestern Europe	South and east across Europe	Protection from anemia
CFTR	Cystic fibrosis	Southeast Europe/Middle East	West and north across Europe	Protection from diarrhea
HbS	Sickle cell disease	Africa/Middle East	To New World	Protection from malaria
FV Leiden	Blood clots	Western Europe	Worldwide	Protection from sepsis
ALDH2	Alcohol toxicity	Far East Asia	North and west across Asia	Protection from alcoholism, possibly hepatitis B
LCT	Lactose tolerance	Asia	West and north across Eurasia	Allows consumption of milk from domesticated animals
GJB2	Deafness	Middle East	West and north across Europe	Unknown



Seventy Percent of All Cystic Fibrosis Alleles Are Due to One Deletion Because of Founder Effect



Factor VIII Mutations Occur Throughout the Gene and are Family Specific

[*Haemophilia* 11, 481-491 (2005)]

Factor VIII gene mutations in haemophilia A patients without intron 22 inversion.

VIII:C (%)	Family history	Consanguinity*	Inversion	Codon†	Mutation	Amino acid change	Exon	Conservation‡
1	Sporadic	NC	Normal	51	TTT → TCT§	Phe → Ser	2	FFFF, identical
1.20	Sporadic	NC	Normal	80	GTT → GAT	Val → Asp	3	VVVV, identical
1	Sporadic	NC	Normal	102	GGT → GTT§	Gly → Val	3	GGGG, identical
2	Sporadic	NC	Normal	104	TCC → CCC§	Ser → Pro	3	SSSS, identical
6	Sporadic	NC	Normal	143	GAG → AAG§	Glu → Lys	4	EEEE, identical
1	Sporadic	NC	Normal	233	delCA§	Thr → fs (TGA-264)	6	
2.70	Inherited	NC	Normal	321	GAA → AAA	Glu → Lys	8	EEEE, identical
0	Sporadic	NC	Normal	372	CGC → CAC	Arg → His	8	RRRR, identical
3	Inherited	NC	Normal	527	CGG → TGG	Arg → Trp	11	RRRR, identical
1	Sporadic	NC	Normal	528	TGC → TAC§	Cys → Tyr	11	CCCC, identical
1	Inherited	NC	Normal	592	CAA → TAA	Gln → Stop	12	QQQQ, identical
1	Inherited	NC	Normal	864	delGACA insCAATTAAATGAGAA§	Gly → fs [TAA-867]	14	
1	Sporadic	NC	Normal	948	insA§	Lys → fs (TGA-984)	14	
1	Sporadic	NC	Intron 1	1107	AGG → TGG§	Arg → Trp	14	RGKK, dissimilar
1	Sporadic	NC	Normal	1107	AGG → TGG§	Arg → Trp	14	RGKK, dissimilar
1	Inherited	NC	Normal	1191–1194	delA	Ile → fs (TAG-1198)	14	
1.40	Sporadic	NC	Normal	1191–1194	insA	Ile → fs (TAA-1220)	14	
1	Sporadic	C	Normal	1227	delC§	Leu → fs (TGA-1231)	14	
2.10	Sporadic	NC	Normal	1241	GAC → GAG	Asp → Glu	14	DGGE, similar
1	Sporadic	NC	Normal	1392	1392del1418§	Pro → fs (TAG-1446)	14	
1	Inerited	C	Normal	1392	1392del1418§	Pro → fs (TAG-1446)	14	
1	Sporadic	NC	Normal	1441	insA§		14	
1	Inerited	C	Normal	1441	insA§			
1	Inherited	NC	Normal	1502	CAG → TAG§	Gln → Stop	14	QREQ, dissimilar
1	Inherited	NC	Normal	1504	delGT§	Val → fs (TGA-1517)	14	
1	Sporadic	NC	Normal	1535	TGG → TGA	Trp → Stop	14	WLWM, dissimilar
nhibitor 96 BU								
1	Sporadic	NC	Normal	1571	TAT → TAA§	Tyr → Stop	14	Y-YY, dissimilar
1	Sporadic	NC	Normal	1581	AAA → TAA§	Lys → Stop	14	KEKK, dissimilar
0.20	Sporadic	NC	Normal	1696	CGA → GGA	Arg → Gly	14	RRRR, identical
1.80	Sporadic	NC	Normal	1729	delA§	Gln → fs (TAA-1752)	15	
1	Inherited	NC	Normal	1751	GAA → AAA§	Glu → Lys	15	EEEE, identical
1	Sporadic	NC	Normal	1775	TTC → TCC§	Phe → Pro	16	FFFF, identical
1	Sporadic	NC	Normal	1835	TGG → TGA§	Trp → Stop	16	WWWW, identical
7.60	Sporadic	C	Normal	1882	ATC → ATA§	Ile → Ile	17	IIII, identical
3	Inherited	C	Normal	1966	CGA → CAA	Arg → Glu	18	RRRR, identical
1	Sporadic	NC	Normal	1966	CGA → TGA	Arg → Stop	18	RRRR, identical

FVIII GENE MUTATIONS IN INDIAN PATI

Need To Screen Across the Gene for Markers -- Family Specific

DNA Testing Should Be Carried Out On Every Individual Born in the US:

a. Yes
b. No

DNA Testing Results Should Be Made Widely Available?

a. Yes
b. No

SNPs Can Be Associated/Linked With Specific Traits & Used By Genetic Testing Companies

OCA2

From SNPedia

OCA2, the oculocutaneous albinism gene (also known as the human P protein gene, or, DN10), is a gene associated with albinism and certain pigmentation effects in general such as eye color, skin color, and hair color.

A large (>3,000 individuals) study of Caucasians indicates that the following **OCA2** variants, all located in the first intron of the gene, are preferentially linked to blue eye color inheritance; together, they form haplotypes that (in some cases at least) predict eye color with greater than 50:50 odds. [PMID 17236130; OMIM 203200.0013 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0013)]

- rs7495174
- rs6497268
- rs11855019

The haplotypes are defined in order as listed above for these 3 SNPs, so, for example, the TGT haplotype refers to rs745174(T)-rs6497268(G)-rs11855019(T). The correspondence between diplotypes (the two haplotypes in one individual) and the % of individuals with blue/gray, green/hazel/ and brown eye color , respectively, was reported as follows for the most common diplotypes[PMID 17236130]:

- TGT/TGT: 62.5, 28.0, 9.5
- TGT/***TTC***: 47.1, 20.3, 32.6
- TGT/***CGT***: 28.6, 14.3, 57.1
- TGT/***TGC***: 27.9, 22.1, 50.0
- ***TGC***/***TTC***: 25.0, 8.3, 66.7
- ***TTT***/***TGC***: 20.7, 31.0, 48.3
- TGT/***TTT***: 17.6, 38.5, 44.0
- TGT/***CTC***: 7.9, 23.3, 68.8

The haplotypes shown in ***bold italics*** represent the ones reported by the authors of this study to be most associated with brown eye color. Furthermore, the haplotypes shown above are as published, and the associated SNPs - which have since changed # as well - are not in the orientation shown in dbSNP.

More recently, a study of a large Danish family led to associations with 2 SNPs in a different region of **OCA2** as linked to blue or brown eye color:

- rs12913832
- rs1129038

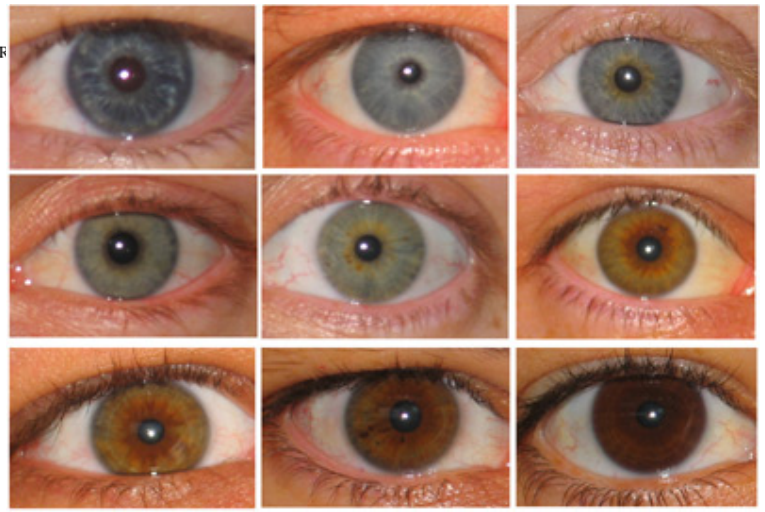
Earlier studies found different regions of the **OCA2** gene to also be predictive of eye color;

- **OCA2** SNP rs1800401 helps predict brown eye color. [PMID 12163334, PMID 15889046; OMIM 203200.0011 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0011)]
- **OCA2** SNP rs1800407 may be associated with green/hazel eye color in some populations, but not others. [PMID 12163334, PMID 15889046; OMIM 203200.0012 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0012)]

is a	gene
is	mentioned by
wikipedia	OCA2 (http://en.wikipedia.org/wiki/OCA2)
google	OCA2 (http://www.google.com/search?hl=en&q=OCA2)
gopubmed	OCA2 (http://www.gopubmed.org/search?q=OCA2)
23andMe	OCA2 (https://www.23andme.com/you/explorer/gene/?gene_name=OCA2)
GeneRIF	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=4948&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_R
dbSNP	4948 (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=4948&chooseRs=all)
PubMed	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Link&LinkName=gene_pubmed&from_uid=4948)
HugeNav	4948 (http://hugenavigator.net/HuGENavigator/huGEPedia.do?firstQuery=OCA2}&geneID=4948&typeSubmit=GO&check=y&typeOption=gene&which=2&pubOrderType=pubD)

	Chromosome position
Rs1129038	26,030,454
Rs11631797	26,175,874
Rs12593929	26,032,853
Rs1800401	25,933,648
Rs1800407	25,903,913
Rs2238289	26,126,810
Rs2240203	26,167,797
Rs28934272	25,903,842
Rs3935591	26,047,607
Rs3940272	26,142,318
Rs4778241	26,012,308
Rs7170852	26,101,581
Rs7183877	26,039,328
Rs7495174	26,017,833
Rs8028689	26,162,483
Rs916977	26,186,959

SNPs in Human P Protein (OCA2) Gene Lead To Different Eye Colors (Physical & Molecular Markers)



Human Eye Color

Using Large Populations SNPs Can Be Used As Markers For Specific Genes/ Traits

SNPedia (http://www.snpedia.com/)

- New model for prostate cancer based on 5 SNPs
- rs1815739 sprinters vs endurance athletes
- rs4420638 and rs429358 can raise the risk of Alzheimer's disease by more than 10x
- rs6152 can prevent baldness
- rs9939609 triggers obesity
- rs662799 prevents weight gain from high fat diets
- rs7495174 green eye color
- rs7903146 in 3% of the population greatly increases the risk of type-2 diabetes
- rs12255372 linked to type-2 diabetes and breast cancer
- rs2395029 asymptomatic HIV viral load set point
- rs324650 influences intelligence and alcohol dependence
- rs1799971 makes alcohol cravings stronger
- rs17822931 determines earwax

Caution

How Will You Use the Information?
How Good Are The Correlations?
What To Do With The Information?
Privacy Issues?
Group Differences? Discrimination?

Constructing Portraits From DNA

Research & Discovery

Portrait in DNA

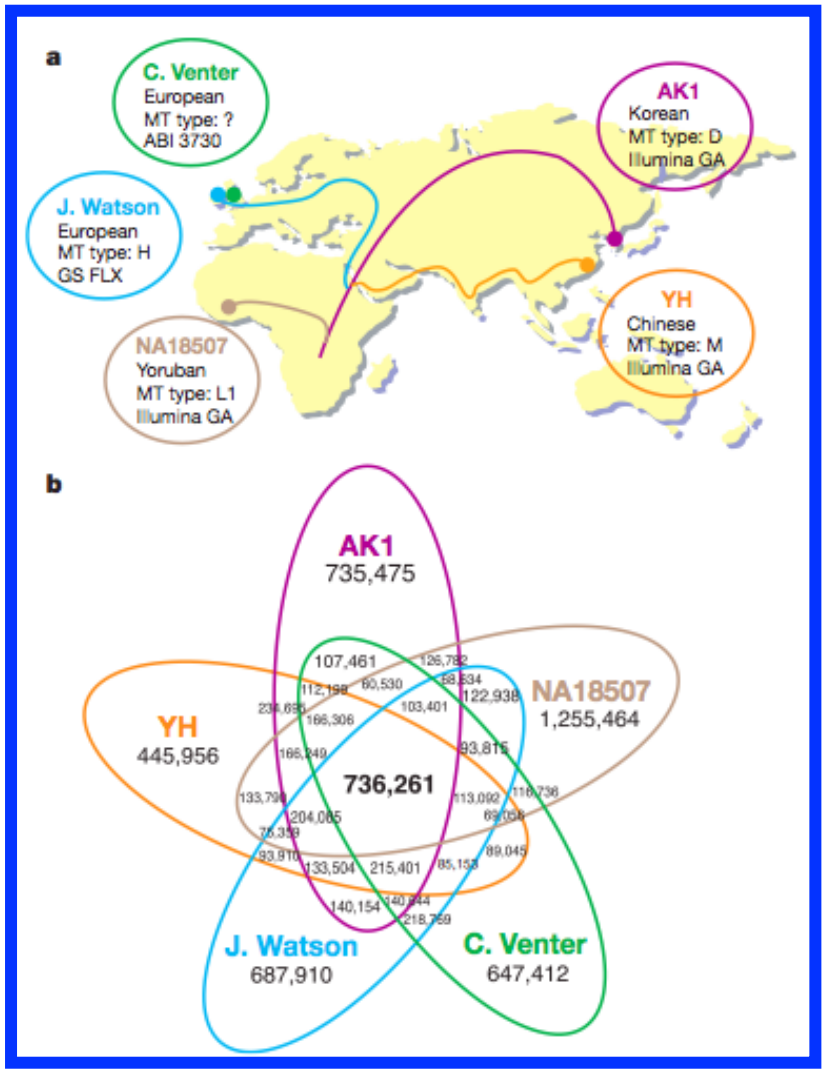
Can forensic analysis yield police-style sketches of suspects? BY CHRISTINE SOARES

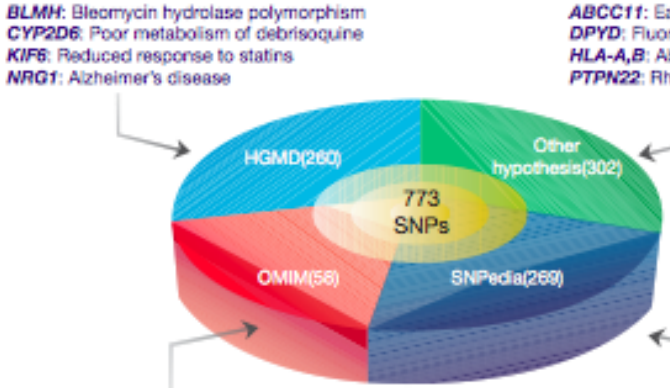
MALE, SHORT AND STOUT, WITH DARK SKIN, BROWN EYES, shovel-shaped teeth, type A+ blood and coarse, dark brown hair giving way to pattern baldness. He would have a high tolerance for alcohol and a higher-than-average risk of nicotine dependence—fortunately, he lived thousands of years before humans discovered smoking. The description of a Stone Age Greenland resident published in February paints an extraordinary portrait of a man who vanished more than 4,000 years ago, drawn almost solely from his DNA remains.

RECONSTRUCTED: Ancient DNA provided details about the looks of a man who lived n Greenland more than 4,000 years ago.

A highly annotated whole-genome sequence of Korean individual

Nature, August 20, 2009, 460, 1011-1016





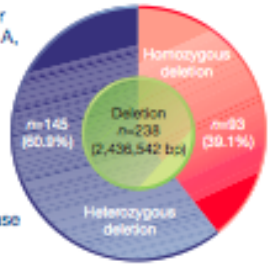
BLMH: Bleomycin hydrolase polymorphism
CYP2D6: Poor metabolism of debrisoquine
KIF6: Reduced response to statins
NRG1: Alzheimer's disease

ABCC11: Earwax related
DPYD: Fluorouracil toxicity
HLA-A,B: Abacavir hypersensitivity
PTPN22: Rheumatoid arthritis

ADD1: Salt-sensitive hypertension
ADRB2: Asthma, reduced β-2 agonist efficacy
LOXL1: Glaucoma
MMP9: Lumbar disc herniation
MTHFD1: Neural tube defects
NDUFV2: Parkinson's disease
SLC30A8: NIDDM
SP110: TB infection
UGT1A1: Gilbert's syndrome

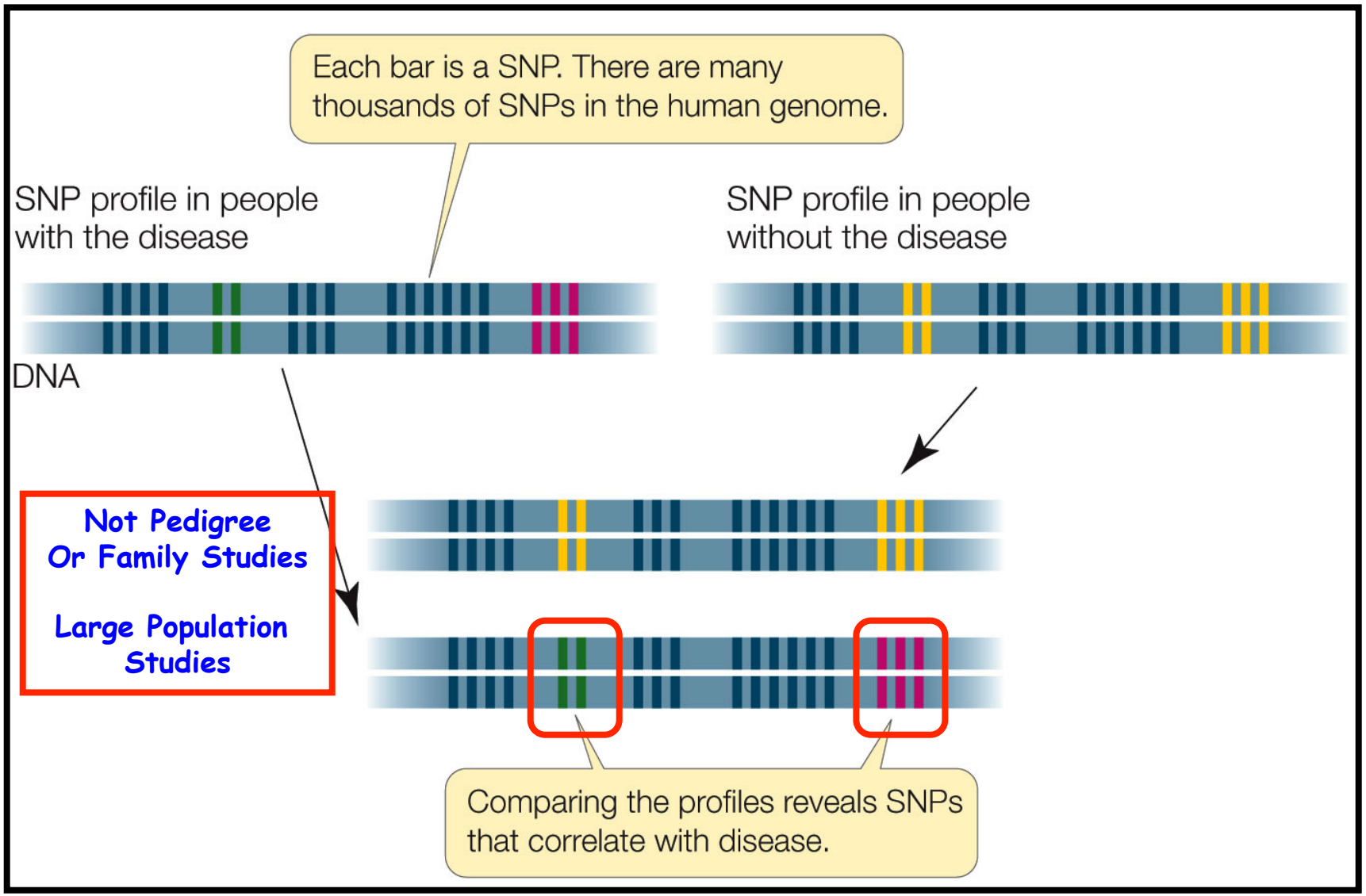
Higher type II DM risk
Higher prostate cancer risk
Better rosiglitazone efficacy
Warfarin resistance
Better citalopram response

- ***ACOT1***: Acyl-CoA thioesterase 1
- ***ANGPT1***: Angiopoietin 1
- ***APBA1***: Amyloid β A4 precursor protein-binding, family A, member 1
- ***CARD6***: Caspase recruitment domain-containing protein 16
- ***FREM2***: Fras1-related extracellular matrix protein 2
- ***GCH1***: GTP cyclohydrolase 1
- ***GSTT2***: Glutathione S-transferase theta 2
- ***SGK1***: Serum/glucocorticoid-regulated kinase 1

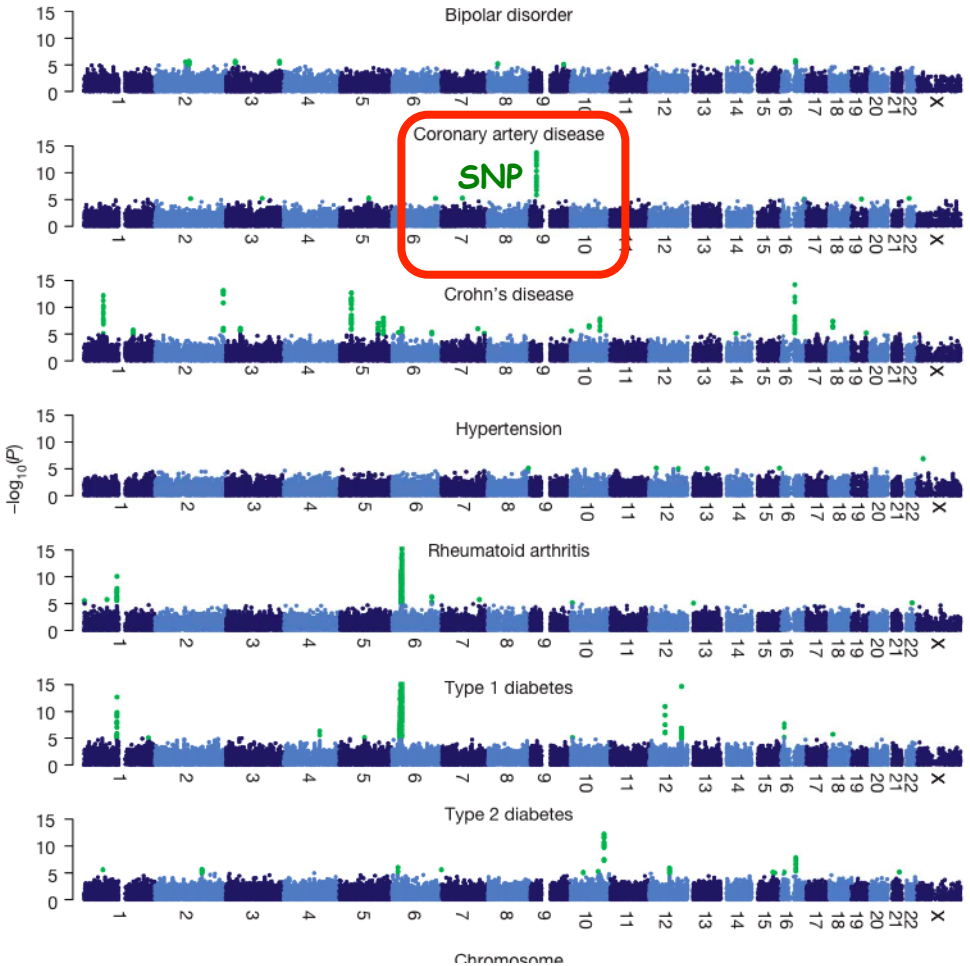


- ***CAMTA1***: Calmodulin-binding transcription activator 1
- ***FRAS1***: Fraser syndrome 1
- ***LILRA3***: Leukocyte immunoglobulin-like receptor
- ***PADI4***: Peptidylarginine deiminase
- ***TP63***: Tumour protein p63
- ***TPH2***: Tryptophan hydroxylase 2
- ***TRY6***: Trypsinogen C

Using SNPs and Population Association Studies to Find Disease Markers and Genes



Correlating SNPs With Specific Diseases Using SNP Chips & Association Studies



Note: Probability

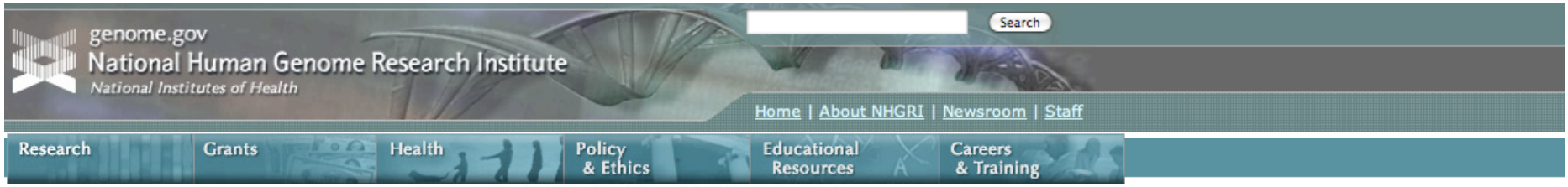
SNPs May Be Near Or In Relevant Genes

TABLE 17.5

SNP Human Genome Scans and Diseases

DISEASE	LOCATION OF SNP (CHROMOSOME NUMBER)	% INCREASED RISK	
		HETEROZYGOTES	HOMOZYGOTES
Breast cancer	8	20	63
Coronary heart disease	9	20	56
Heart attack	9	25	64
Obesity	16	32	67
Diabetes	10	65	277
Prostate cancer	8	26	58

All Published Genome-Wide Association Studies Are Listed on the National Human Genome Research Institute Website



Correlate SNPs With Specific Traits And Used By Personal Gene Testing Companies Such as 23andMe®

SNPs Become Personal!

It's All About Me

Along with the flood of discoveries in human genetics, 2007 saw the birth of a new industry: personal genomics. Depending on your budget, you can either buy a rough scan of your genome or have the whole thing sequenced. The companies say the information will help customers learn about themselves and improve their health. But researchers worry that these services open up a Pandora's box of ethical issues.

At $300,000 to $1 million per genome, sequencing all 3 billion base pairs is still too costly for all but a few. Although dozens more personal genomes will probably be sequenced in the coming year, most will be done by public and private research organizations—including the institute run by genome maverick J. Craig Venter, whose personal genome was one of three completed in 2007 in the United States and China. In a lower-budget effort, Harvard's George Church this month will deliver initial DNA sequences for the protein-coding sections (1% of the genome) to the first 10 volunteers for his Personal Genome Project. Meanwhile, a new company called Knome is offering full-genome sequencing to 20 customers willing to pay $350,000.

A glimpse of one's genome is already within the reach of ordinary people, thanks to several companies. They include 23andMe, which has financing from Google and may let users link to others with shared traits; Navigenics, which will screen for about 20 medical conditions; and deCODE Genetics in Iceland, a pioneer in disease gene hunting. For $1000 to $2500, these companies will have consumers send in a saliva sample or cheek swab, then use "SNP chips" to scan their DNA for as many as 1 million markers. The companies will then match the results with the latest publications on traits, common diseases, and ancestry.

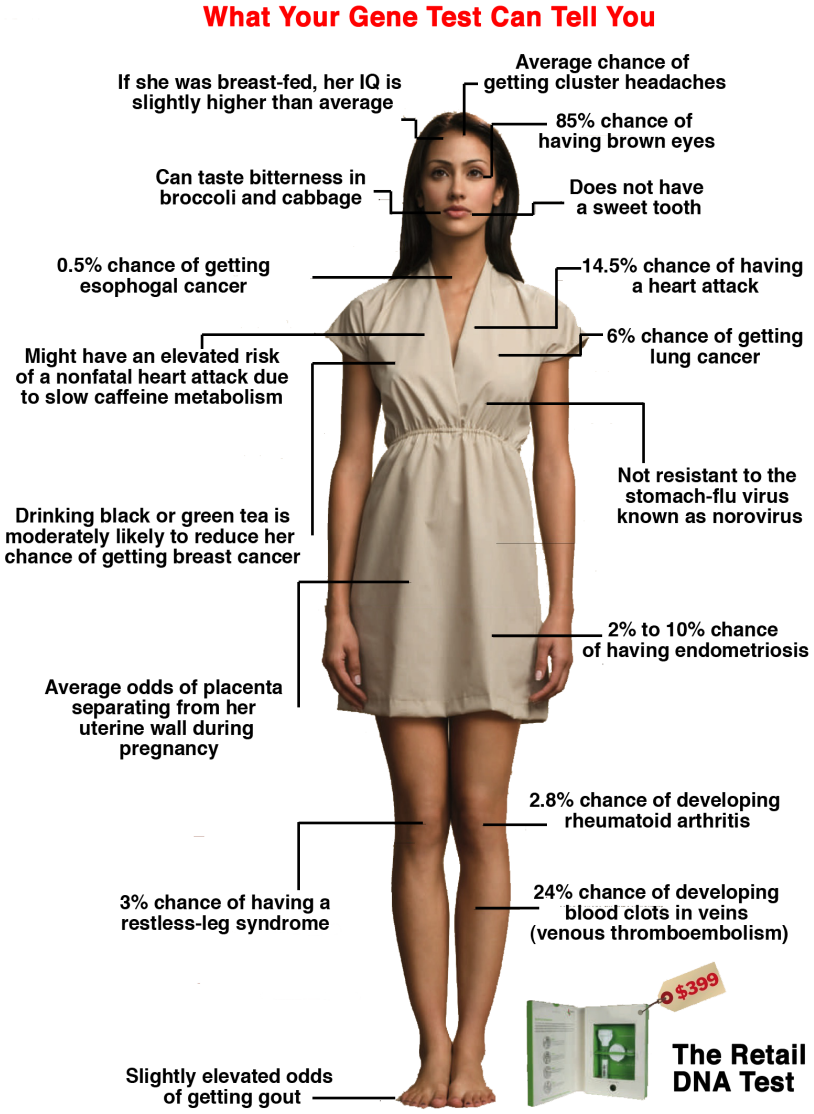
Although many customers may view this exercise as a way to learn fun facts about themselves—recreational genomics, some call it—bioethicists are wary. Most common disease markers identified so far raise risks only slightly, but they could cause needless worry. At the same time, some people may be terrified to learn they have a relatively high risk for an incurable disease such as Alzheimer's.

Pandora's box? This cheek-swab kit could reveal your intimate secrets.

The rush toward personal genome sequences also sharpens long-held worries about discrimination. A bill to prevent insurers and employers from misusing genetic data is stalled in Congress. Complicating matters, your genetic information exposes your relatives' DNA, too.

The most profound implications of having one's genome analyzed may not be what it reveals now—which isn't much—but what it may show later on. Perhaps to sidestep such questions, some companies will limit which markers to disclose. Others, however, will hand customers their entire genetic identity, along with all the secrets it may hold.

–JOCELYN KAISER



Invention Of the Year

The Problems With Human Genome Testing Companies Are

a. Reliability of Results?
b. Privacy?
c. What To Do With Information Obtained?
d. Regulatory Oversight?
e. All of Above?

Problem: Different Companies-Different Predictions!

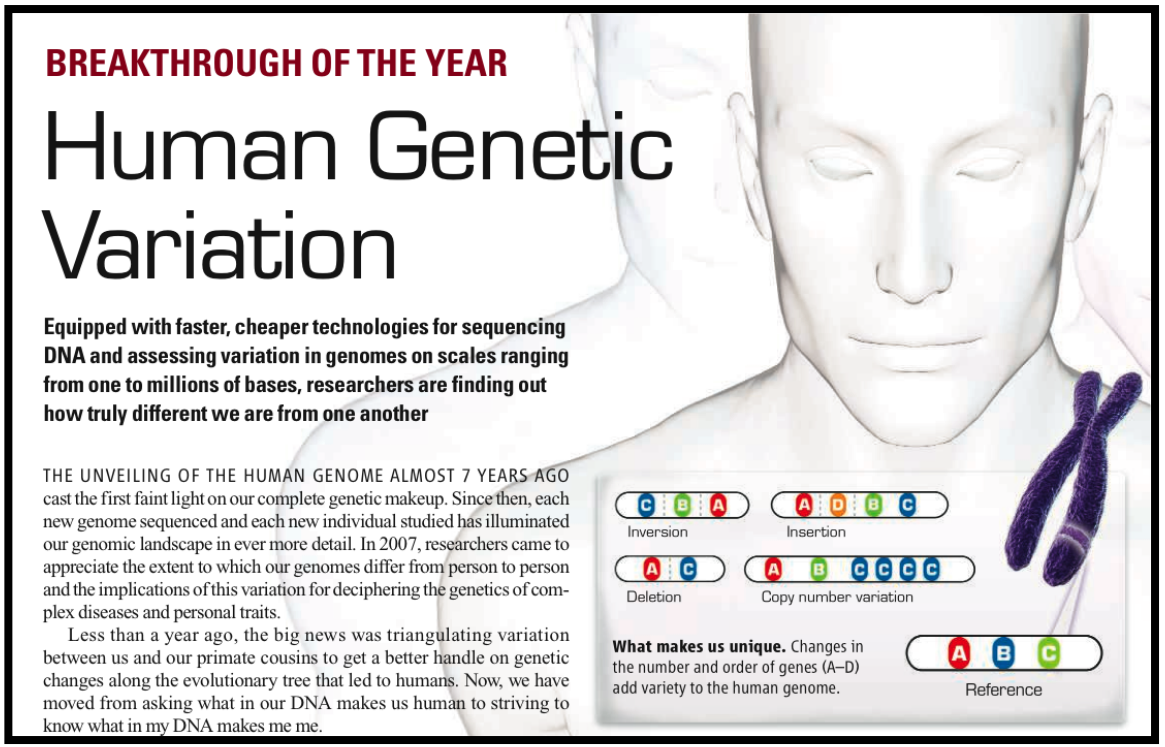
TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS

Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk (RR > 1.05), ↓ decreased risk (relative risk (RR) < 0.95), = average risk (0.95 ≤ RR ≤ 1.05). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

Identifying DNA Variations Between Individuals Has Other Uses

1. Marking and Identifying Disease Genes
2. Paternity, Individual Identification, Forensics
3. Human Population History and Origins



BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.

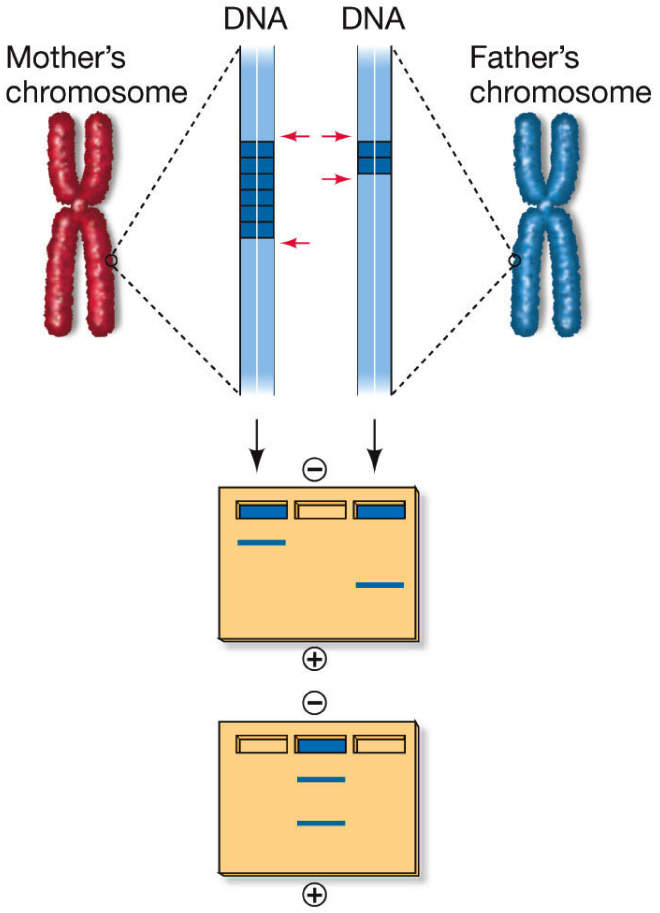
What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

STRs (VNTRs) Used to Verify Remains of Russian Royal Family

	Number of repeats	
STR-1	15,16	15,16
STR-2	8,8	7,10
STR-3	3,5	7,7
STR-4	12,13	12,12
STR-5	32,36	11,32

Tsarina Alexandra ○ □ Tsar Nicholas II

STR-1	15,16	15,16	15,16
STR-2	8,10	7,8	8,10
STR-3	5,7	5,7	3,7
STR-4	12,13	12,13	12,13
STR-5	11,32	11,36	32,36



PNAS

Genomic identification in the historical case of the Nicholas II royal family

PNAS, March, 2009

Evgeny I. Rogaev[a,b,c,d,1], Anastasia P. Grigorenko[b,d], Yuri K. Moliaka[b], Gulnaz Faskhutdinova[b], Andrey Goltsov[d], Arlene Lahti[e], Curtis Hildebrandt[e], Ellen L. W. Kittler[f], and Irina Morozova[a]

[a]Department of Genomics and Laboratory of Evolutionary Genomics, Vavilov Institute of General Genetics, Russian Academy of Science, Gubkina Street, 3, Moscow, 119991, Russian Federation; [b]Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, 303 Belmont Street, Worcester, MA 01604; [c]Faculty of Bioinformatics and Bioengineering, Lomonosov Moscow State University, Moscow, 119991, Russian Federation; [d]Research Center of Mental Health, Russian Academy of Medical Science, Zagorodnoe Shosse 2/2, Moscow, 113152, Russia; [e]Molecular World, Inc., Thunder Bay, ON, Canada P7B 2T1; and [f]University of Massachusetts Medical School, Center for AIDS Research, Worcester, MA 01605

Communicated by James D. Watson, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, November 14, 2008 (received for review October 8, 2008)

RESEARCH ARTICLE OPEN ACCESS

Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis

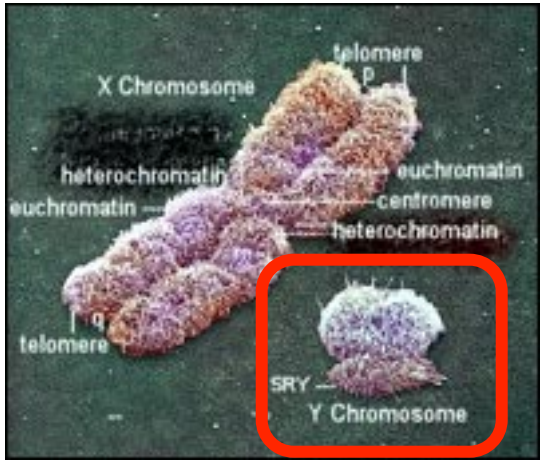
Michael D. Coble[1#*], Odile M. Loreille[1#], Mark J. Wadhams[1], Suni M. Edson[1], Kerry Maynard[1], Carna E. Meyer[1], Harald Niederstätter[2], Cordula Berger[2], Burkhard Berger[2], Anthony B. Falsetti[3], Peter Gill[4,5], Walther Parson[2], Louis N. Finelli[1]

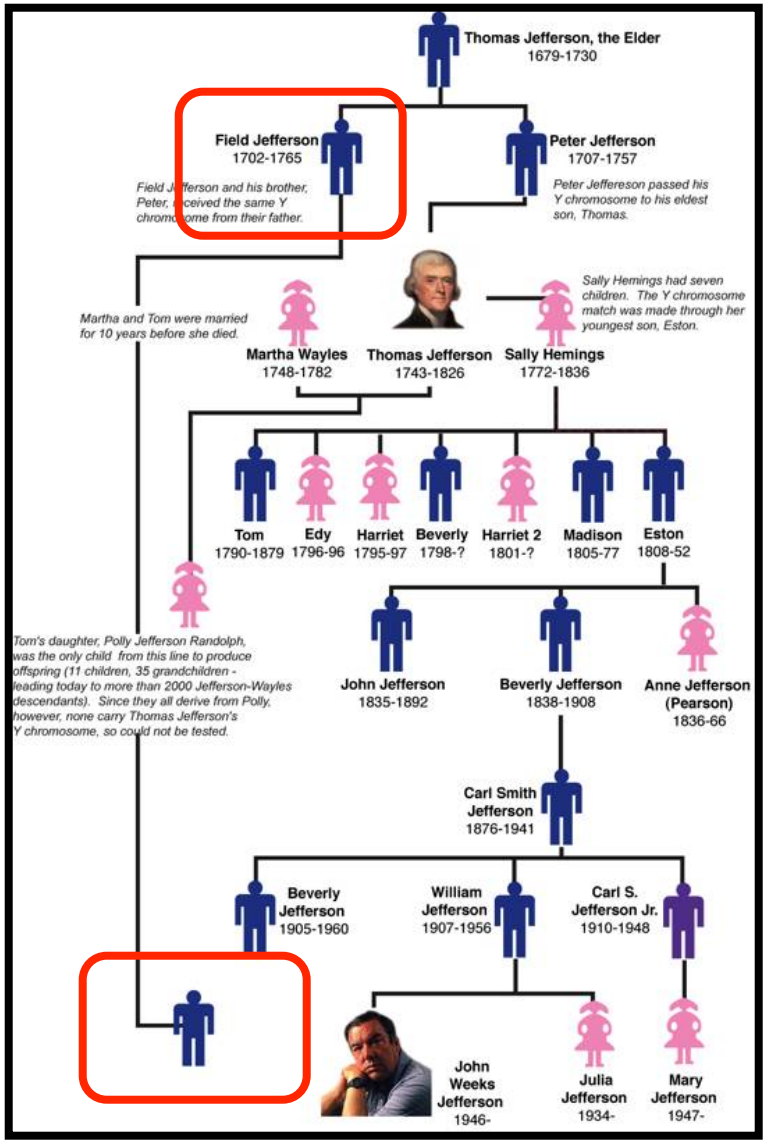
1 Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology, Rockville, Maryland, United States of America, 2 Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria, 3 University of Florida, Gainesville, Florida, United States of America, 4 Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, United Kingdom, 5 Institute of Forensic Medicine, University of Oslo, Oslo, Norway

Download: XML | PDF | Citation
E-mail this Article
Order Reprints
Print this Article
Bookmark this page:

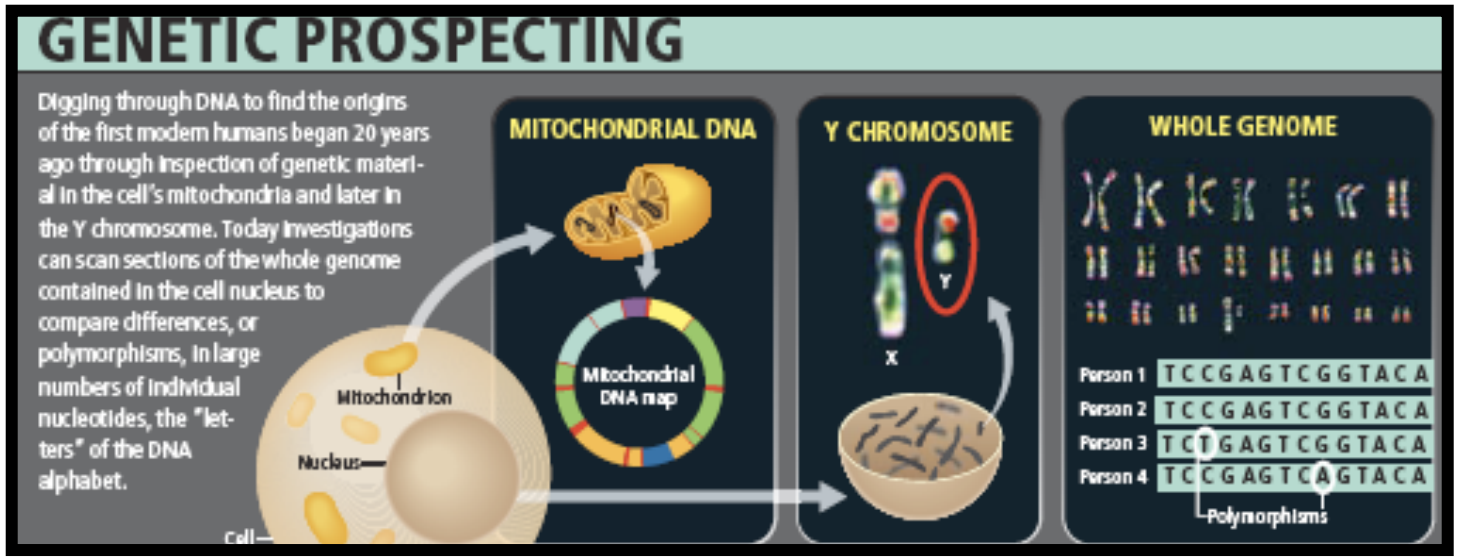
PLOS, March, 2009

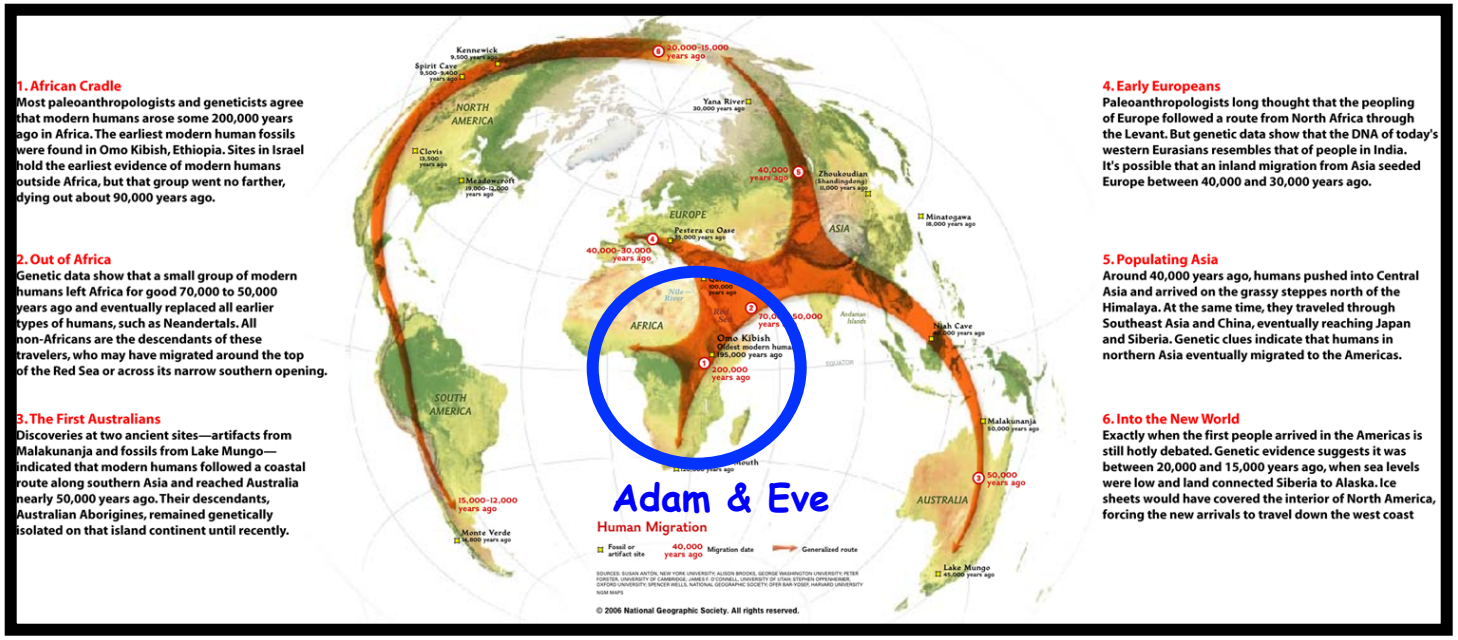
Using Y Chromosome SNPs and RFLPs To Determine That Thomas Jefferson and Sally Hemmings Had Children



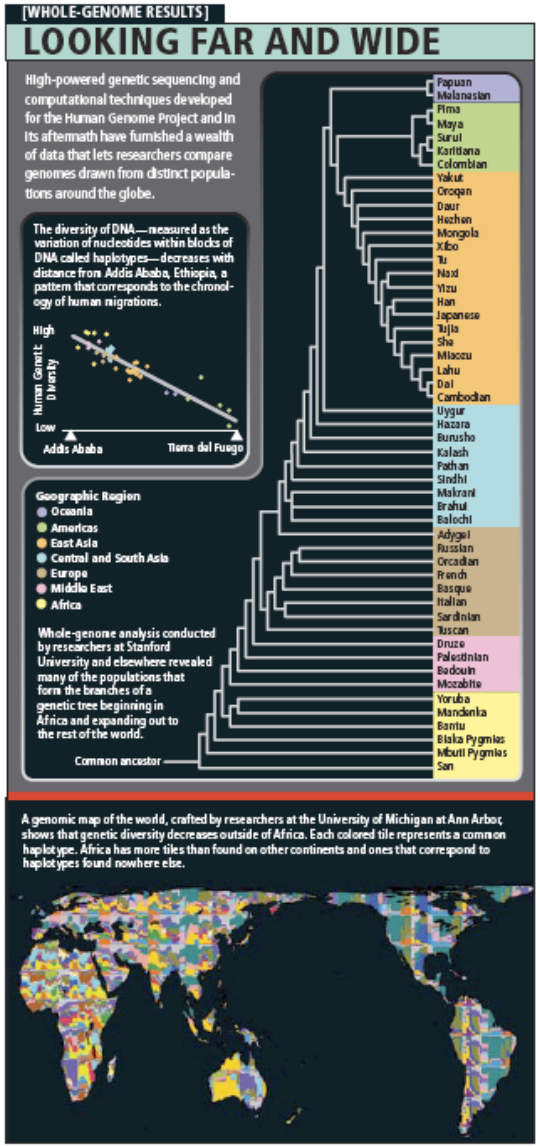


Tracing Human Populations Using SNPs and DNA Polymorphisms





Human Population Relationships Using Whole-Genome Comparisons



Most Genetic Diversity In African Populations

HUMAN DIVERSITY

Scientific American Library
1982 ISBN 07167-14698

RICHARD LEWONTIN

Human Races Have a Genetic Basis:

a. Yes
b. No

There is More Genetic Diversity Within Populations than Between Populations!! So Much for the Concept of Racial "Purity"!!!!

More Genetic Diversity Within Any Population Than Between Polulations

Proportion of genetic diversity accounted for within and between populations and races

Gene	Total $H_{species}$	Proportion		
		Within Populations	Within Races between Populations	Between Races
Hp	.994	.893	.051	.056
Ag	.994	.834	—	—
Lp	.639	.939	—	—
Xm	.869	.997	—	—
Ap	.989	.927	.062	.011
6PGD	.327	.875	.058	.067
PGM	.758	.942	.033	.025
Ak	.184	.848	.021	.131
Kidd	.977	.741	.211	.048
Duffy	.938	.636	.105	.259
Lewis	.994	.966	.032	.002
Kell	.189	.901	.073	.026
Lutheran	.153	.694	.214	.092
P	1.000	.949	.029	.022
MNS	1.746	.911	.041	.048
Rh	1.900	.674	.073	.253
ABO	1.241	.907	.063	.030
Mean		.854	.083	.063

Source: R. C. Lewontin, *Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

1. 85% of Human Genetic Variations Occurs within Populations & Between Individuals in that Populations!
2. Remaining 15% of Human Genetic Variation Split Between Different Populations of Same "race" (8%) & Between Different "races" (6%)
3. Only 6% of Human Genetic Variation are to Differences between races!!! Mostly Geographic. Note: THERE ARE GROUP DIFFERENCES!

Within Population Differences Account For 95% of Human Genetic Variation

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]

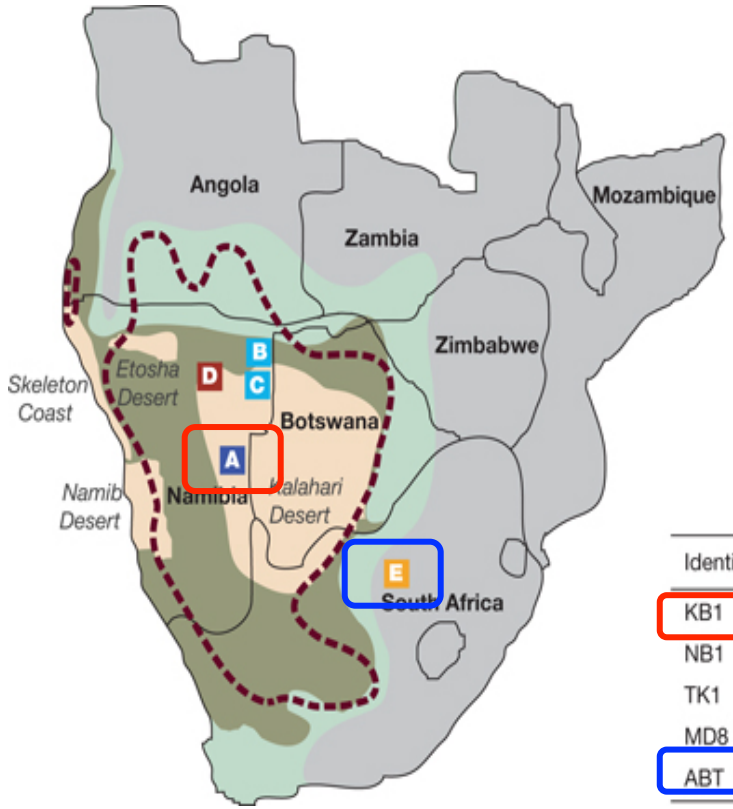
We studied human population structure using genotypes at 377 autosomal microsatellite loci in 1056 individuals from 52 populations. Within-population differences among individuals account for 93 to 95% of genetic variation; differences among major groups constitute only 3 to 5%. Nevertheless, without 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.

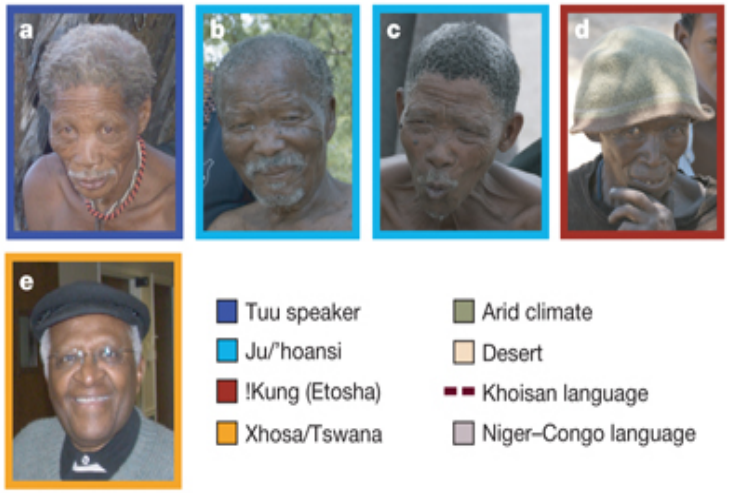
Table 1. Analysis of molecular variance (AMOVA). Eurasia, which encompasses Europe, the Middle East, and Central/South Asia, is treated as one region in the five-region AMOVA but is subdivided in the seven-region design. The World-B97 sample mimics a previous study (*6*).

Sample	Number of regions	Number of populations	Variance components and 95% confidence intervals (%)		
			Within populations	Among populations within regions	Among regions
World	1	52	94.6 (94.3, 94.8)	5.4 (5.2, 5.7)	
World	5	52	93.2 (92.9, 93.5)	2.5 (2.4, 2.6)	4.3 (4.0, 4.7)
World	7	52	94.1 (93.8, 94.3)	2.4 (2.3, 2.5)	3.6 (3.3, 3.9)
World-B97	5	14	89.8 (89.3, 90.2)	5.0 (4.8, 5.3)	5.2 (4.7, 5.7)
Africa	1	6	96.9 (96.7, 97.1)	3.1 (2.9, 3.3)	
Eurasia	1	21	98.5 (98.4, 98.6)	1.5 (1.4, 1.6)	
Eurasia	3	21	98.3 (98.2, 98.4)	1.2 (1.1, 1.3)	0.5 (0.4, 0.6)
Europe	1	8	99.3 (99.1, 99.4)	0.7 (0.6, 0.9)	
Middle East	1	4	98.7 (98.6, 98.8)	1.3 (1.2, 1.4)	
Central/South Asia	1	9	98.6 (98.5, 98.8)	1.4 (1.2, 1.5)	
East Asia	1	18	98.7 (98.6, 98.9)	1.3 (1.1, 1.4)	
Oceania	1	2	93.6 (92.8, 94.3)	6.4 (5.7, 7.2)	
America	1	5	88.4 (87.7, 89.0)	11.6 (11.0, 12.3)	

But - There Are Differences! But…They Fall Into Geographical Groups -- Groups Divided Originally by Geographic Barriers (Ocean, Desert, Mountains). The 5% Difference Allows Us to Mark and Trace Ancestry!

Recent Sequencing of Two African Genomes Reveals Remarkable Genetic Diversity





Identifier	Name	Location of origin	Linguistic group	Y chromosome
KB1	!Gubi	Southern Kalahari	Tuu-speaker	B2b
NB1	G/aq'o	Northern Kalahari	Juu (Ju/'hoansi)	A3b1
TK1	D#kgao	Northern Kalahari	Juu (Ju/'hoansi)	A2
MD8	!Aî	Northern Kalahari	Juu (!Kung)	E1b1b1
ABT	Tutu	South Africa	Bantu (Xhosa/Tswana)	E1b1a8a *

SC Schuster et al. Nature 463, 943-947 (2010)

Each Genome Contains One Million SNPs Not Found in Any Other Genome

Conclusions

1. If 85% of Human Genetic Variation Occurs Between Different People Within Any Given Population (localized)
2. If only 7% of Human Genetic Variation Occurs Between "Races" (novel alleles specific to "races") e.g. F_{yB} ^{ES}
3. Then Losing all "Races" Except One Retains 94% of all Human Genetic Variation!

[85% + (15%-7%)]= 94%

Variation That Occurs in Ancestral Population

85% Within Population genetic variability

8% Between Populations of Same "Race"

7% Between "Race" Genetic Variability

4. ∴ Human Highly Heterozygous or Hybrids- & If Above Not True- Most of Us Would Not Be Here- Need Genetic Variation to Survive!

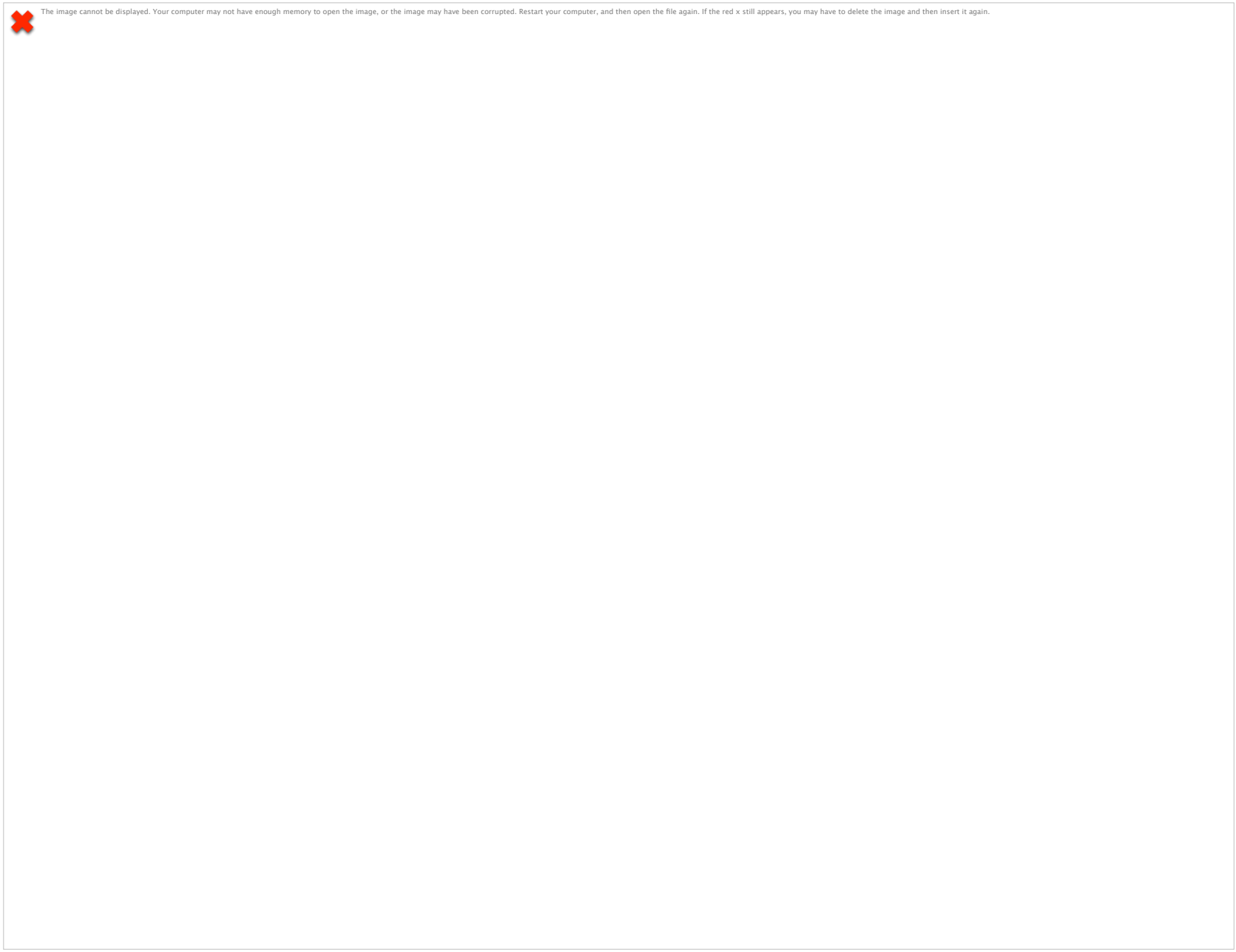
So What is a "Race"?

1. Primarily a sociological concept- but could be a localized or inbred population that has a higher frequency of alleles at a very small number of loci. Affects few physical features.
2. High frequency alleles in one "race" are present at lower frequencies in other "races". All humans have same genes- differ in form mostly within populations!
3. Heterozygosity (variation) high in human populations- all populations. None homozygous at all loci!
4. No such thing as a "pure" race - would have little variation
5. Genes affecting physical features not representation of genes across genome-" selected" traits

Geographical Ancestry is relevant-many "racial" groups now have multiple ancestries because of admixture and migration

A Better Term is POPULATIONS!


Knowledge or Certainty: The Ascent of Man Series
Jacob Bronowski, 1973


The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.