

The Human Genome

Medha Bhagwat

National Center for Biotechnology Information
National Institutes of Health

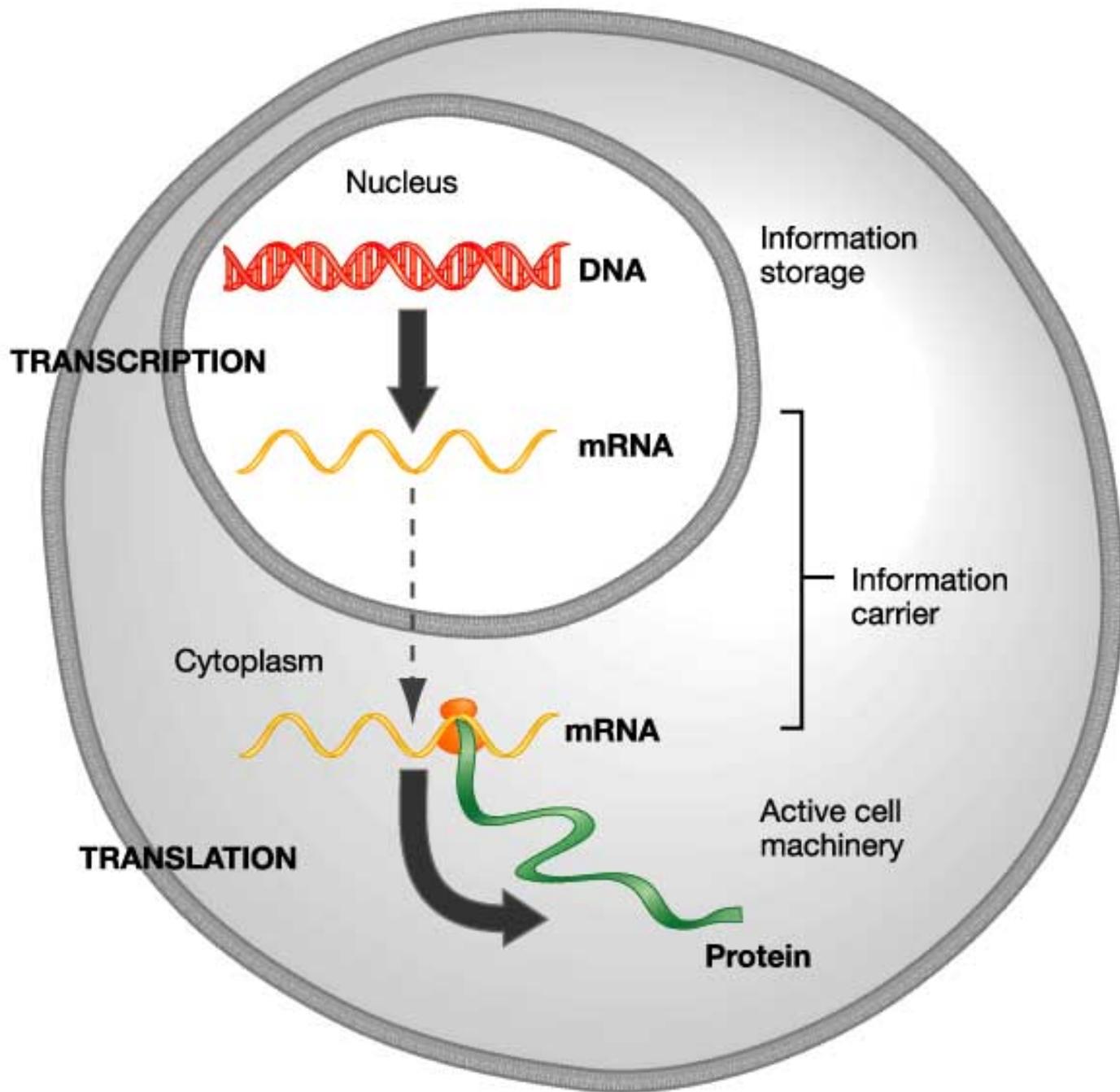


NCBI

Outline

1. What is a genome?
2. Analysis of the human genome
3. How to access the genome data?
4. Future implications
5. Live demo of the NCBI resources

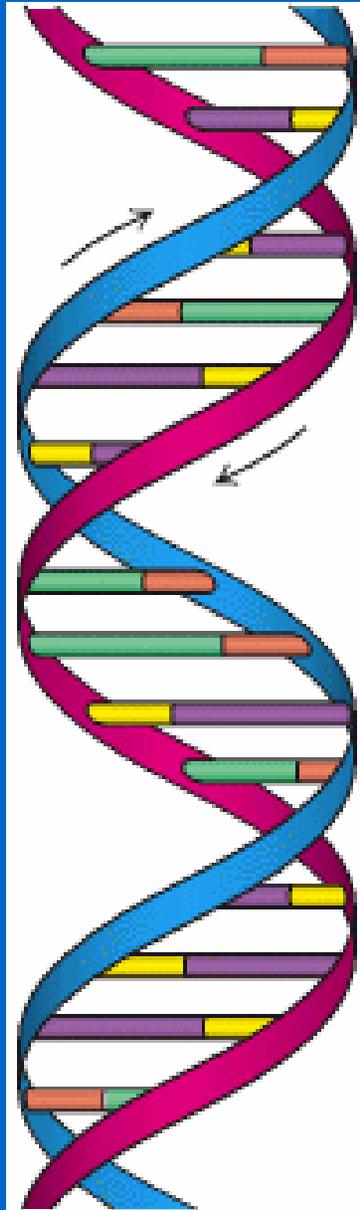
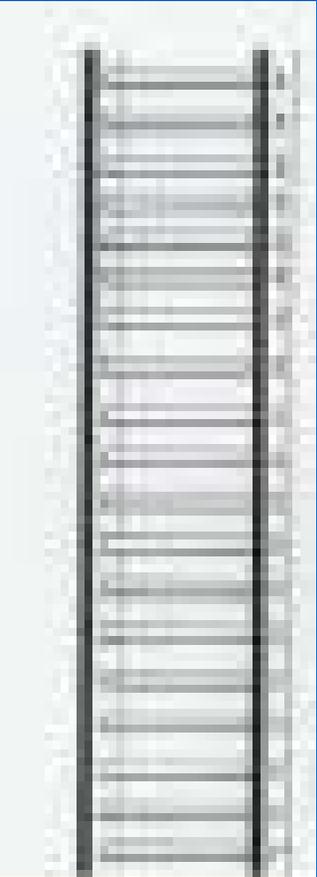




Proteins are Body's Worker Molecules

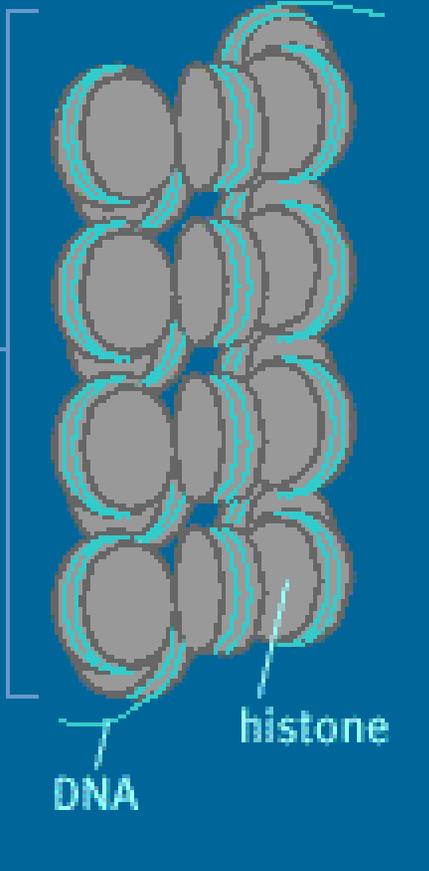
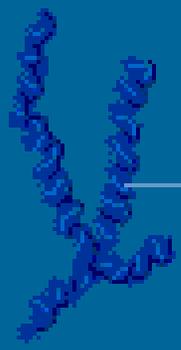


The double-helical structure of DNA



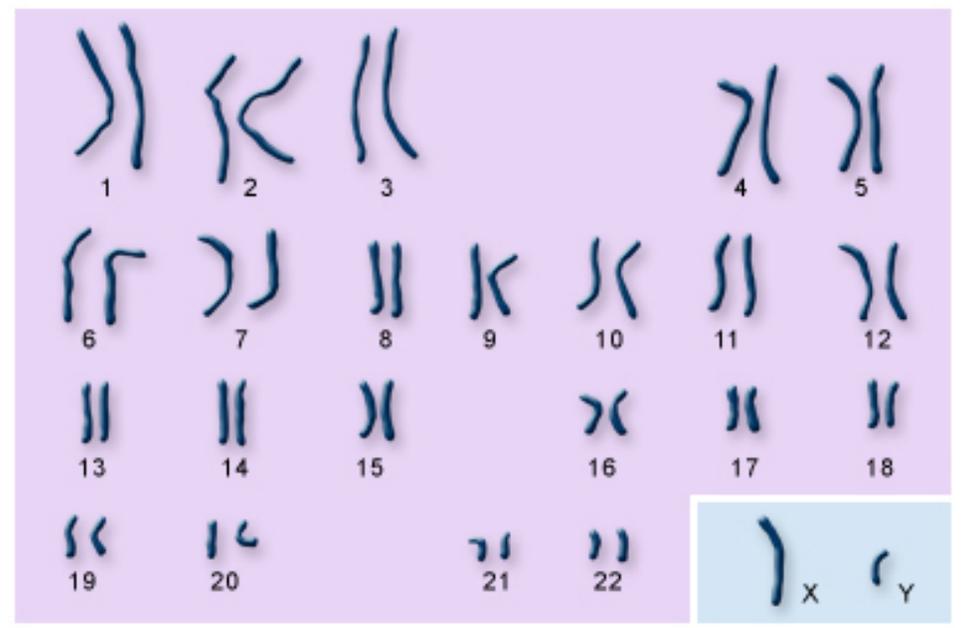
DNA Tower in Perth

chromosome



DNA

histone



autosomes

sex chromosomes

Protein

Alphabet: amino acids

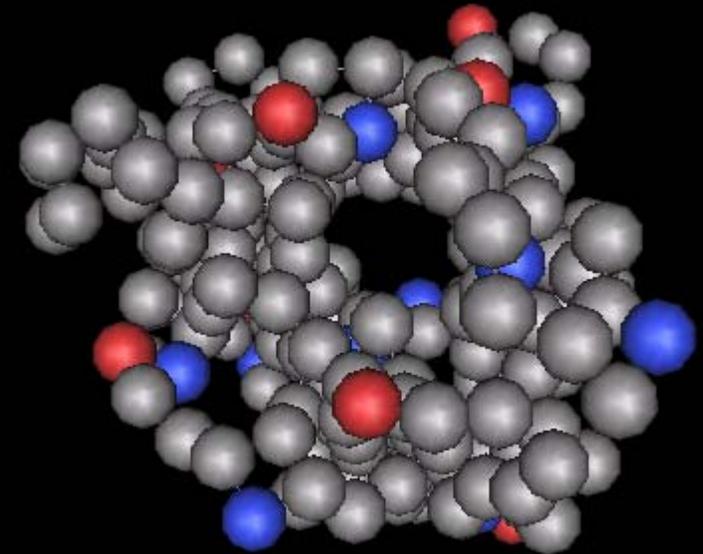
There are 20 amino acids

Encoded by codons (triplets of nucleotides)

ATGTGCAGCCTAGCTGCCGTC

Met—Cys—Ser—Leu —Ala — Ala —Val

Water channel protein



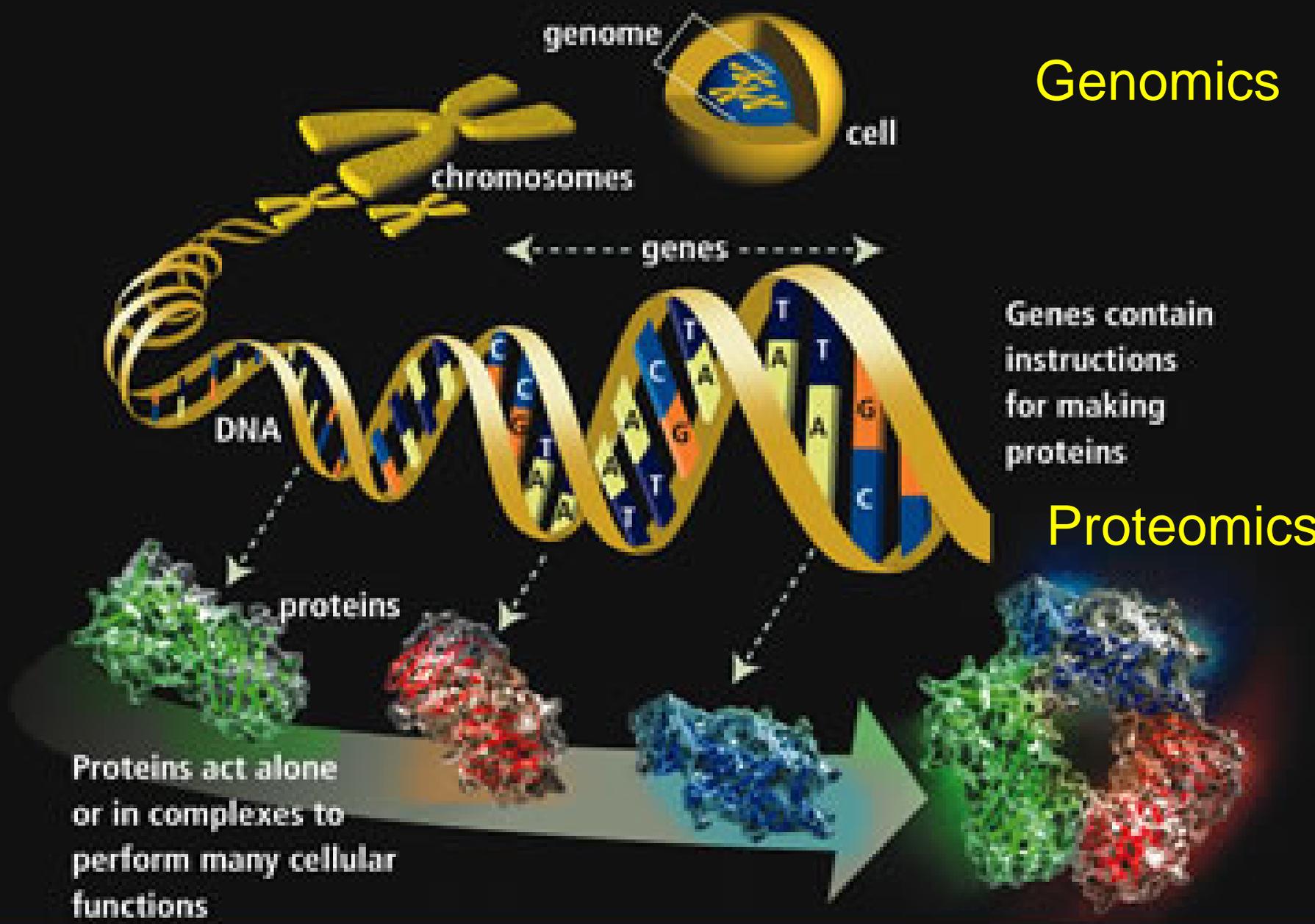
Genomics

Genes contain instructions for making proteins

Proteomics

Proteins act alone or in complexes to perform many cellular functions

From Genes to Proteins



Applications of Genomics and Proteomics

1. Understand basic biology
2. Diagnosis and treatment of diseases
3. Rationale for drug design
4. Protect plant life
5. Understand bacterial resistance
6. Solve environmental problems
7. Develop new energy sources
8. Improve industrial processes
9. Study evolutionary changes among organisms

Outline

1. What is a genome?
2. Analysis of the human genome
3. How to access the genome data?
4. Future Implications
5. Live demo of the NCBI resources



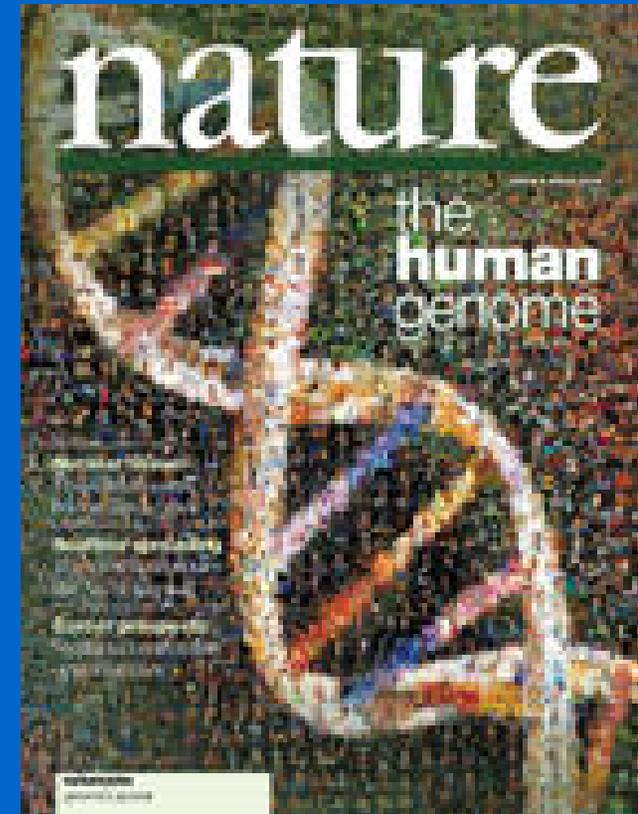
Human Genome Sequenced

Celera



June 23, 2000

The Human Genome Project



The Human Genome

23 pairs of chromosomes

3.2 billion base pairs

Estimated number of genes about 30,000

Only 2% of the human genome “codes”

Average gene size 4000 base pairs

Largest gene dystrophin 2.4 million base pairs

More than 50% in repeat elements or
so called “junk DNA”

Analysis of the Human Genome

The DNA sequence of any two people is 99.9 percent identical.

Sites in the DNA sequence where individuals differ at a single DNA base are called single nucleotide polymorphisms (SNPs).

The SNPs may greatly affect an individual's disease risk.



Genome (DNA)

Exact spelling of a word is necessary

CAT

RAT

MAT

FAT

BAT

EAT

HAT

DAT

GAT

KAT

CBT

CCT

CAQ

CAC



Hemochromatosis

- High absorption of iron
- Resulting in excessive storage of iron, particularly in the liver, skin and pancreas
- a codon change TGT to TAT in the HFE gene
- changes amino acid from cysteine to tyrosine leading to hemochromatosis

[3-D structure of the protein](#)



NCBI



Sickle Cell Anemia



- Sickled red blood cells
- Mutation in the HBB gene that codes for hemoglobin
- one nucleotide change in the 7th codon GAG to GTG
- changing glutamic acid to valine
- interaction between valine and the complementary regions on adjacent molecules results in the formation of polymers that aggregate and distort the shape of the red blood cells

[3-D structure of hemoglobin](#)

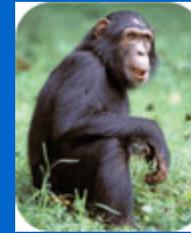
[3-D structure of mutant hemoglobin](#)

Sizes of Different Genomes

Human (<i>Homo sapiens</i>)	3.2 billion
Laboratory mouse (<i>M. musculus</i>)	2.6 billion
Fruit fly (<i>D. melanogaster</i>)	137 million
Yeast (<i>S. cerevisiae</i>)	12.1 million
Bacterium (<i>E. coli</i>)	4.6 million
Human immunodeficiency virus (HIV)	9700

Human Genome Sequenced

What are the genetic changes that make us human?



Only 1.2% difference between Human and chimpanzee:
Every 100th nucleotide different
Could affect thousands of genes
Many are probably the consequence of 6 million years
of genetic drift, with little effect on body or behavior
Other small changes--perhaps in regulatory, noncoding
sequences--may have dramatic consequences

What are the genetic changes that make us human?



Study clinical mutations in the genes that impair key traits and trace the genes' evolution.

For example, FOXP2, the first ever gene associated with human speech.

FOXP2 knock-out mice are squeakless.

A complete understanding of uniquely human traits will, however, include more than DNA such as nurture and nature.

Outline

1. What is a genome?
2. Analysis of the human genome
3. How to access the genome data?
4. Future Implications
5. Live demo of the NCBI resources



National Center for Biotechnology Information

<http://www.ncbi.nlm.nih.gov>

Created as a part of NLM in 1988

- To establish public databases
 - U.S. National DNA Sequence Database
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information



NCBI |

NCBI Databases

Organisms

Genomes (DNA)



mRNA



Protein



NCBI

NCBI Databases



Entrez, The Life Sciences Search Engine

HOME SEARCH SITE MAP

PubMed

Entrez

Human Genome

GenBank

Map Viewer

Search across databases

GO

CLEAR

Help

Welcome to the new Entrez cross-database search page



PubMed: biomedical literature citations and abstracts



PubMed Central: free, full text journal articles



Books: online books



OMIM: Online Mendelian Inheritance in Man



Site Search: NCBI web and FTP sites



Nucleotide: sequence database (GenBank)



Protein: sequence database



Genome: whole genome sequences



Structure: three-dimensional macromolecular structures



Taxonomy: organisms in GenBank



SNP: single nucleotide polymorphism



Gene: gene-centered information



HomoloGene: Eukaryotic homology groups



UniGene: gene-oriented clusters of transcript sequences



CDD: conserved protein domain database



3D Domains: domains from Entrez Structure



UniSTS: markers and mapping data



PopSet: population study data sets



GEO Profiles: expression and molecular abundance profiles



GEO DataSets: experimental sets of GEO data



Cancer Chromosomes: cytogenetic databases



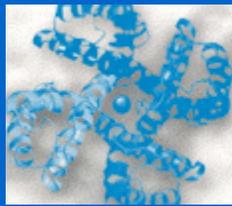
Journals: detailed information about the journals indexed in PubMed and other Entrez databases



MeSH: detailed information about NLM's controlled vocabulary



Bookshelf



BASIC NEUROCHEMISTRY

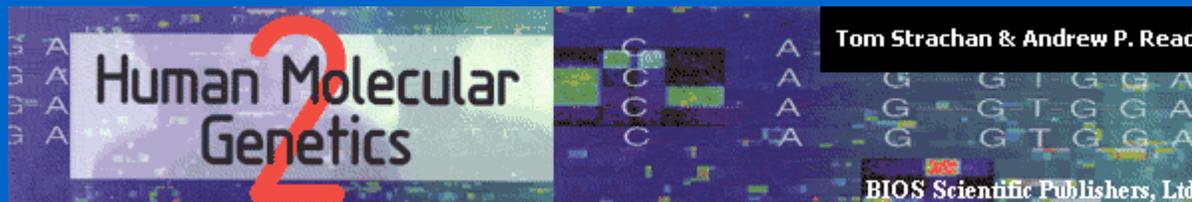
MOLECULAR, CELLULAR
AND MEDICAL ASPECTS
SIXTH EDITION

EDITORS

GEORGE J. SIEGEL
BERNARD W. AGRANOFF
R. WAYNE ALBERS
STEPHEN K. FISHER
MICHAEL D. UHLER



LIPPINCOTT WILLIAMS & WILKINS
A Wolters Kluwer Company



NCBI

Genome Sequence Data and Analysis Tools at NCBI

NCBI

Site Map
guide to NCBI
resources

Cancer
Chromosomes
chromosomal
abnormalities

Clusters of
Orthologous Groups
analysis of
complete
genomes

Gene
gene-related
information

Genome
complete genome
sequences

GEO
gene expression
data

HomoloGene
orthologs

Genomic
Biology

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search Entrez for Go

Genomic Biology

Genomic biology takes a holistic approach to molecular biology and evolution by studying the complete genome, its genes, and its protein expression patterns.

NCBI provides several genomic biology tools and resources, including organism-specific pages that include links to many web sites and databases relevant to that species. We invite you to explore the links provided on this page.

Announcements and Updates

Map Viewer updates:

Species	Build	Release Date
<i>Danio rerio</i> (zebrafish)	Zv4	July 5, 2005 <small>NEW</small>
<i>Anopheles gambiae</i> (mosquito)	2.2	June 30, 2005
<i>Apis mellifera</i> (bee)	2.1	May 31, 2005
<i>Mus musculus</i> (mouse)	34.1	May 19, 2005
<i>Rattus norvegicus</i> (rat)	3.1	April 26, 2005
<i>Bos taurus</i> (cow)	1.1	March 29, 2005

Genome Resources

- Genome Projects Database
 - Eukaryotic
 - Fungi
 - Insects
 - Mammals
 - Microbial
 - Plants
 - Map Viewer
 - Organelles
 - Plant Genomes Center
 - Retroviruses
 - Viral Genomes

Organism-Specific

- Aspergillus G B
- Bee G B
- Cat G B
- Chicken G B

G Genome Resources
B BLAST
M Map Viewer
P Genome Project DB

Outline

1. What is a genome?
2. Analysis of the human genome
3. How to access the genome data?
4. Future Implications
5. Live demo of the NCBI resources



What's Next?????

1. HapMap: Genetic variation mapping project
 - Discovery of genes related to diseases
 - Gene Testing
 - Gene Therapy
2. Pharmacogenomics: Pharmacology and genomics
 - Custom effective drugs based on genetic profile
 - Reduce adverse reactions
3. ENCODE: Encyclopedia of functional elements
 - Study expression of genes



Outline

1. What is a genome?
2. Analysis of the human genome
3. How to access the genome data?
4. Future Implications
5. Live demo of the NCBI resources



Some Useful Web Pages

National Center for Biotechnology Information

<http://www.ncbi.nlm.nih.gov/>

Education

<http://www.ncbi.nlm.nih.gov/Education/>

Science Primer

<http://www.ncbi.nlm.nih.gov/About/primer/>

BookShelf

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>

National Library of Medicine

<http://www.nlm.nih.gov/>

Genetics Home Reference

<http://ghr.nlm.nih.gov/>

