

فناوری اطلاعات و پروژه ژنوم انسان

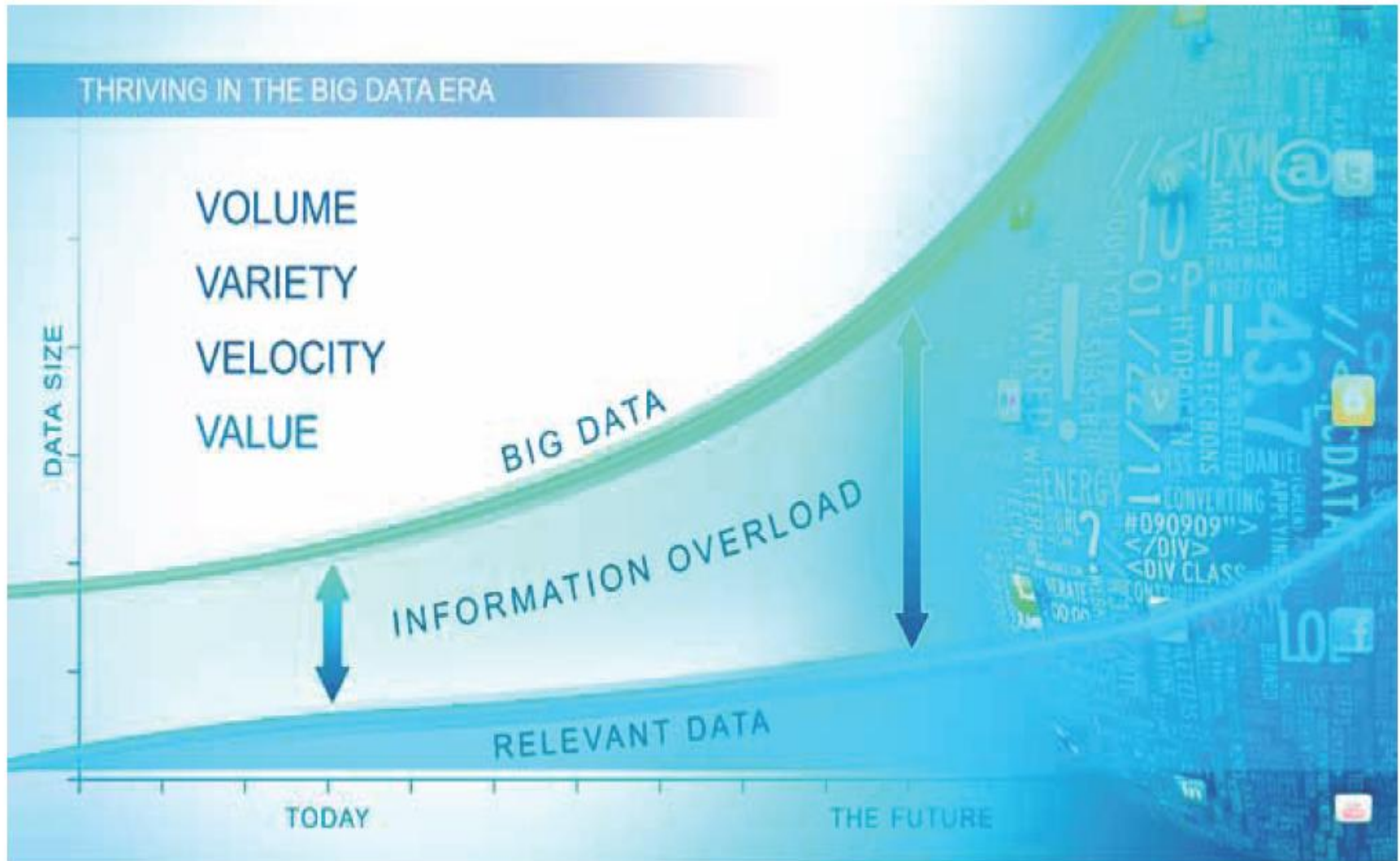
علی متولی زاده اردکانی

پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری

داده های بزرگ (Big Data)



THE BIG DATA ERA



Amount of new data stored varies across geography

New data stored¹ by geography 2010

Petabytes



¹ New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.

SOURCE: IDC storage reports; McKinsey Global Institute analysis

الگوهای پنهان

همبستگی‌های ناشناخته

TOP SECRET//SI//ORCON//NOFORN

Gmail facebook Hotmail Google skype paltalk AOL mail YouTube

 (TS//SI//NF) PRISM Collection Details 

Current Providers

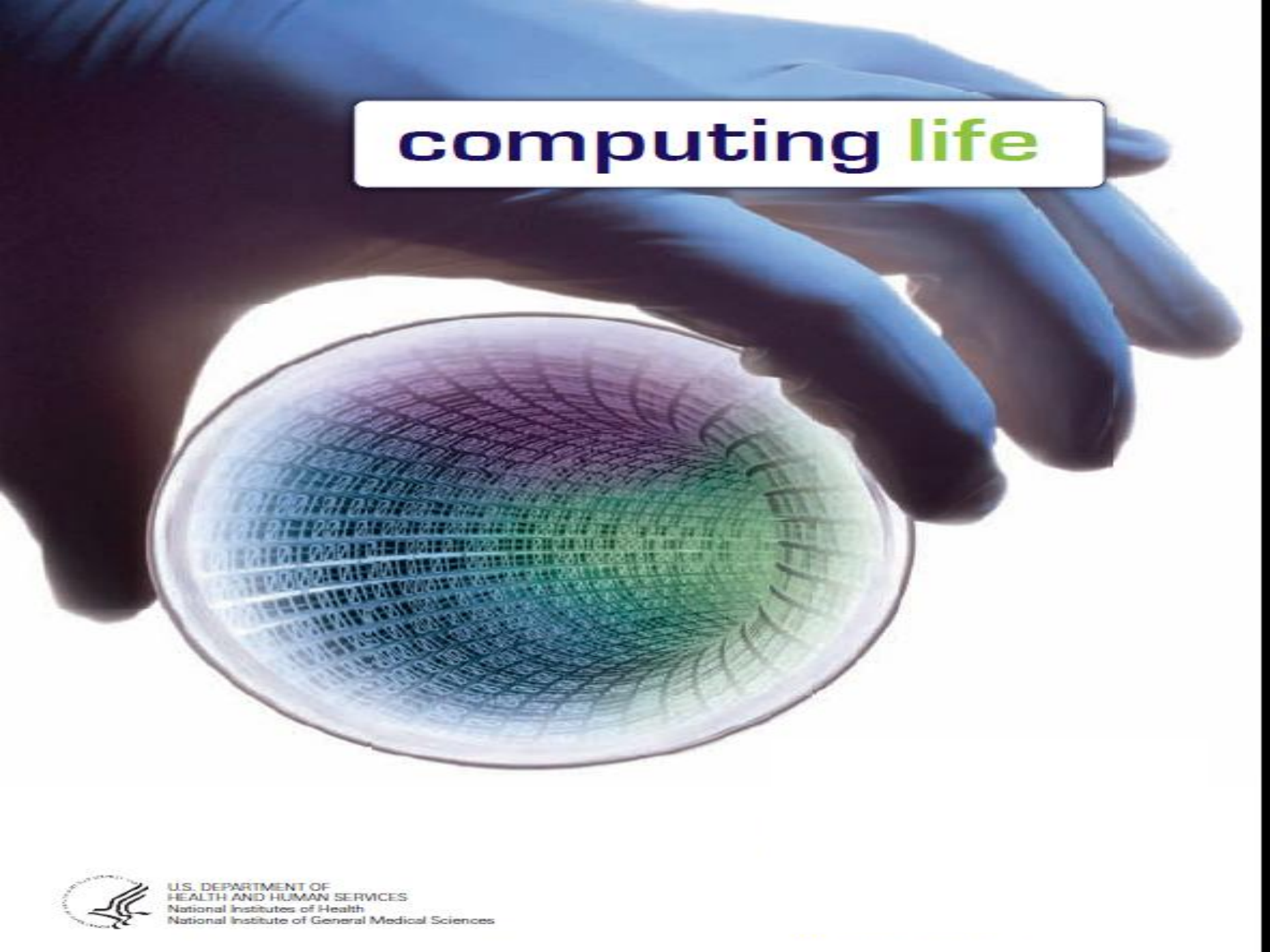
- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple

What Will You Receive in Collection (Surveillance and Stored Comms)?
It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:
Go PRISMFAA

TOP SECRET//SI//ORCON//NOFORN

A blue-gloved hand is shown holding a petri dish. Inside the petri dish, a glowing DNA double helix is visible, surrounded by a pattern of binary code (0s and 1s). The DNA helix is colored with a gradient from purple to green. The binary code is also colored with a gradient from blue to green. The background is white.

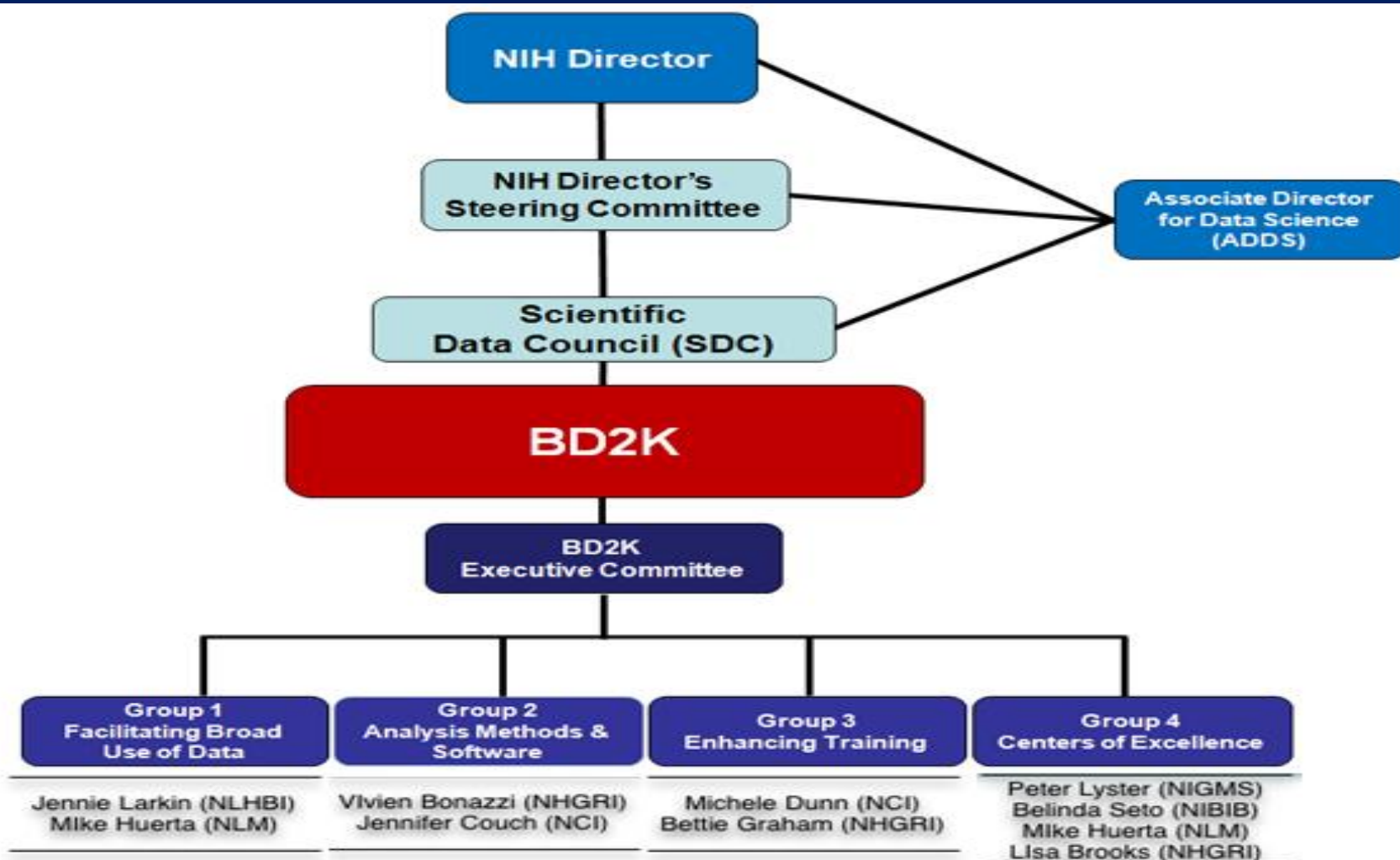
computing life



U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
National Institutes of Health
National Institute of General Medical Sciences

NIH Big Data to Knowledge (BD2K)

2013 - \$96 million to establish six to eight Big Data to Knowledge Centers of Excellence



Opportunities and Challenges in Biomedical Research

- Access
- Management
- Analysis
- Integration of datasets
- [e.g., imaging, phenotypic, molecular (including various '–omics'), exposure, health, behavioral]
- Big Data emanate from three sources:
 - (1) a **small number of groups** that produce very large amounts of data for use by the entire research community;
 - (2) **individual investigators** who produce large datasets, often empowered by the use of readily available **new technologies**
 - (3) research data or **clinical data in electronic health records**

سر فصل ها

- تعریف و ارایه نتایج پروژه ژنوم انسان
- منابع اطلاعات مربوط به پروژه ژنوم انسان
- بیوانفورماتیک و آنالیز اطلاعات مربوط به پروژه ژنوم انسان
- تشخیص زودرس سرطان تخمدان با استفاده از بیوانفورماتیک

پروژه ژنوم انسان



~ 88 years

~50 years

1865

Mendel discovers laws of genetics

1900

Rediscovery of Mendel's genetics

1944

DNA identified as hereditary material

1953

DNA structure

1960's

Genetic code

1977

Advent of DNA sequencing

1975-79

First human genes isolated

1986

DNA sequencing automated

1990

Human genome project officially begins

1995

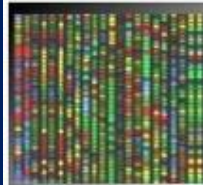
First whole genome

1999

First human chromosome

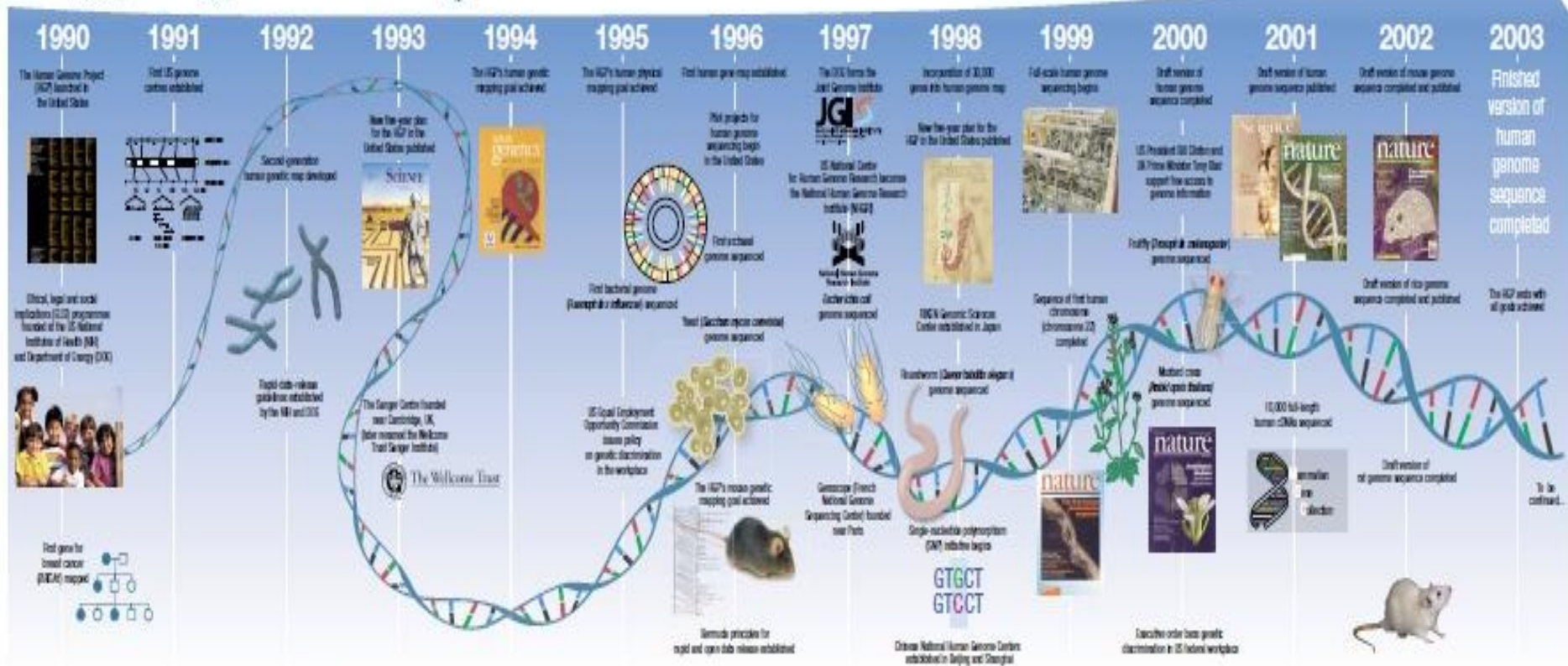
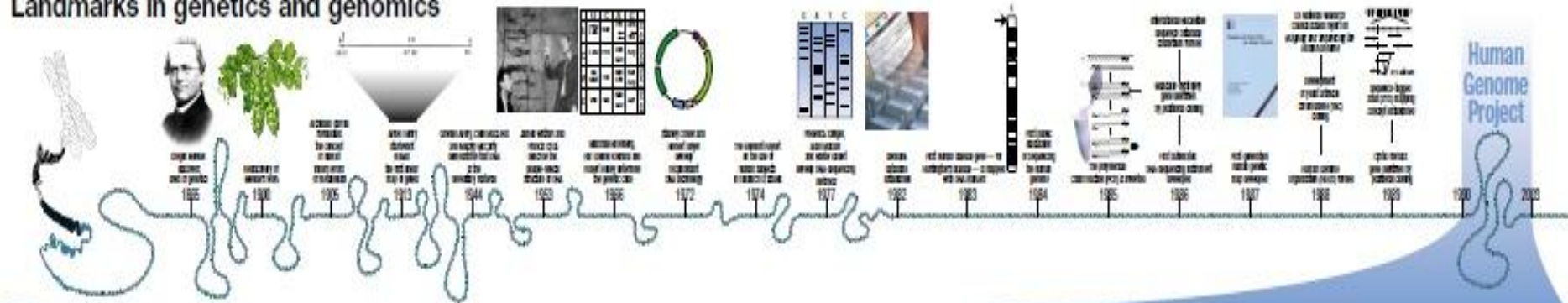
2003

'Finished' human genome sequence



Landmarks in Genetics and Genomics

Landmarks in genetics and genomics



Source: Nature (2003) 421: 20-21. Reproduced with permission of the copyright owner. All rights reserved.

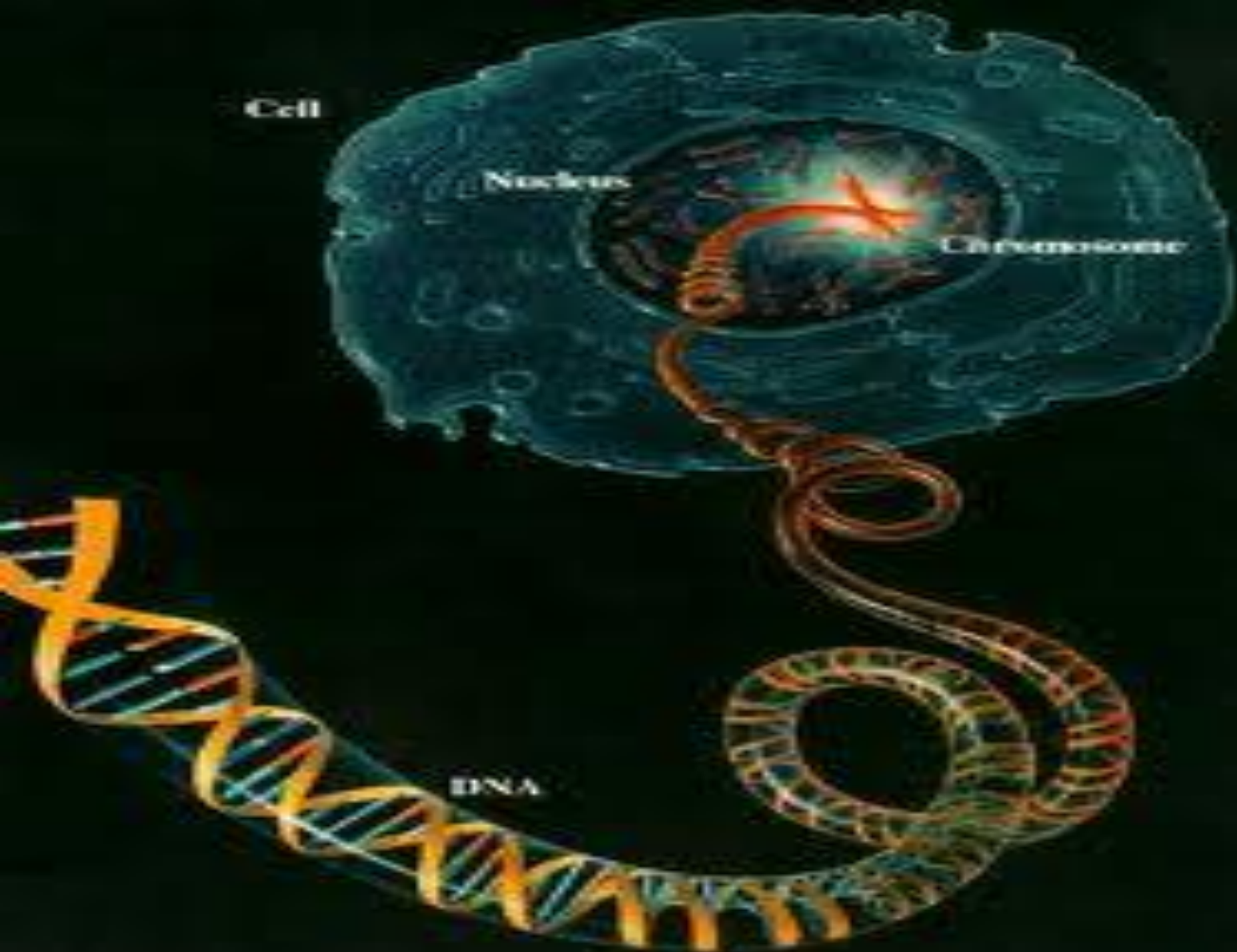
Definitions

- **Genetics:** Study of **Individual genes** and their impact on relatively rare **single gene disorders**
- **Genomics:** Study of **all the genes** in the human genome together, including their **interactions** with **each other**, the **environment**, and other psychosocial and cultural factors

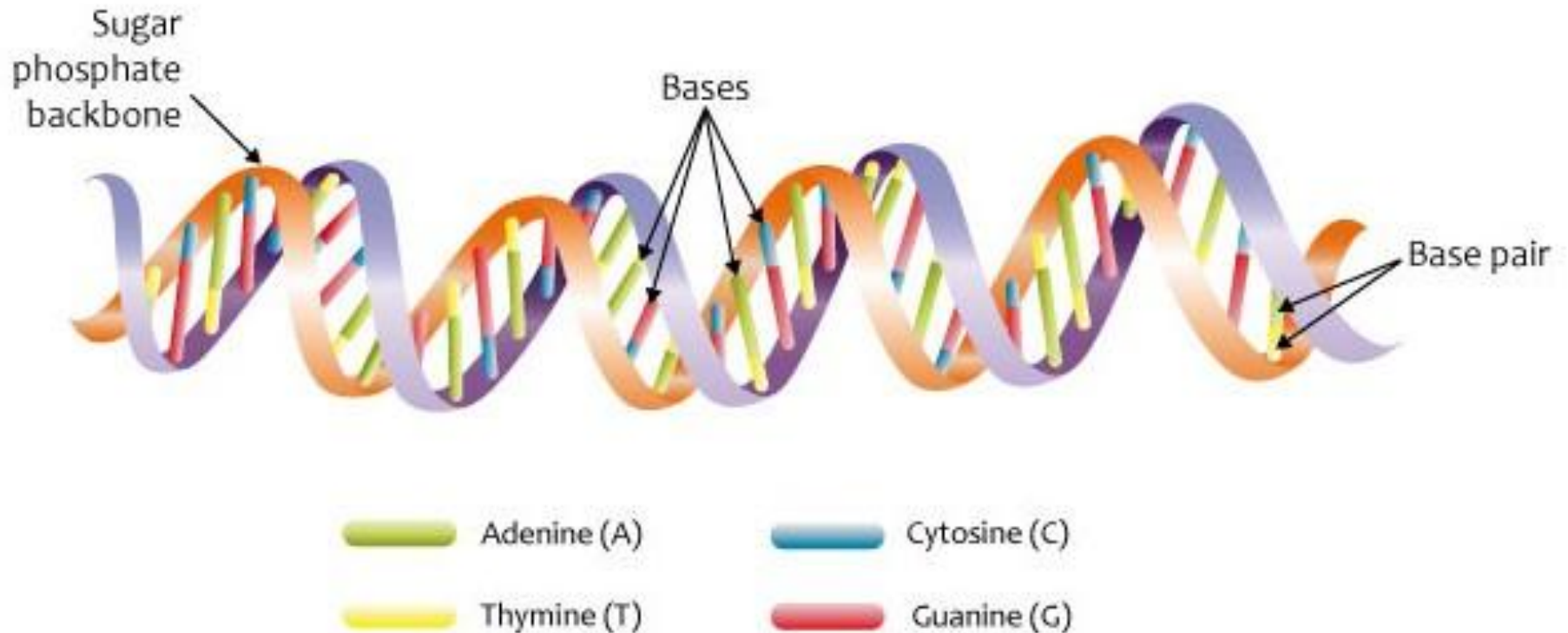
21st Century



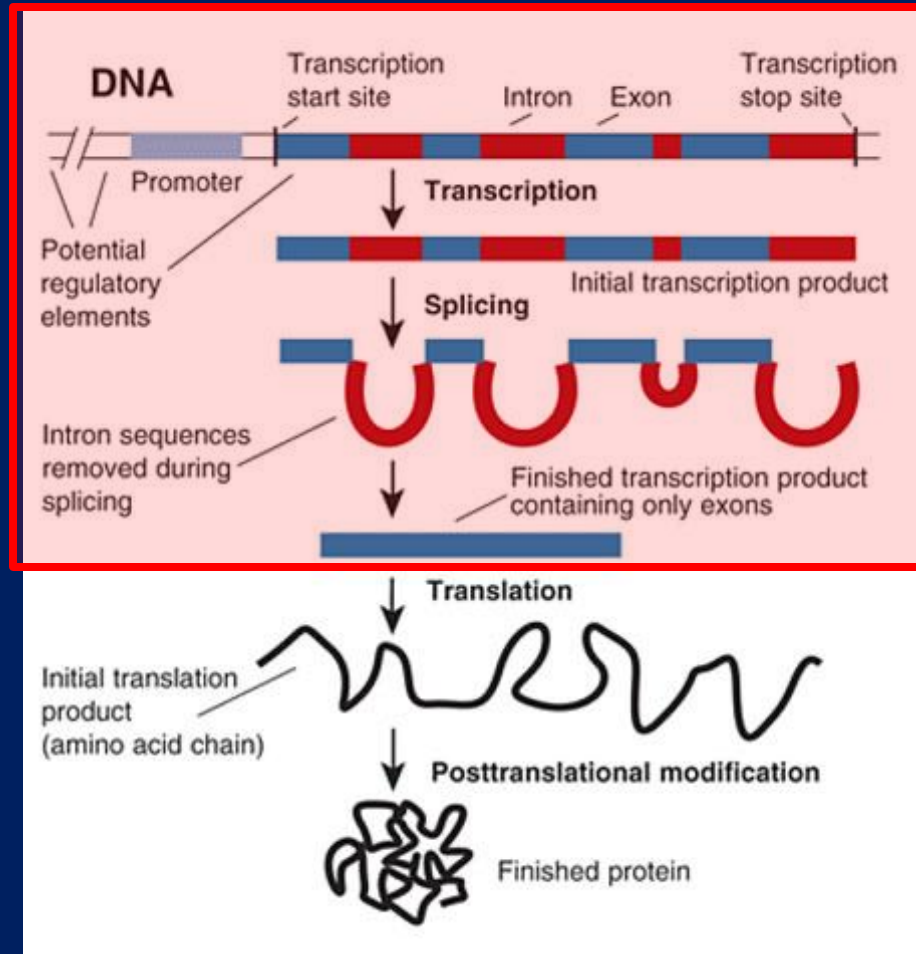
February 2001: First Draft published
April 2003: sequencing is completed but its analysis continue ...



The DNA Double Helix



Gene Expression - Proteins



Human Genome Sequence

10,000 nts x 1000 pg = 300 volumes = 3 Billion bp

AAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGGCCTAAAAGGGGTGCCTAAAAGGGGCCTTGAA
AAGGGCCTTGAAAAGGGCCTTGAAAATTGGHUMANGENOMECCTTGAAAAGGGCCTTGAAAGAGCCTAAAAGGG
GCCTTAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCC
TTGAAAAGCTCCTTGCTAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCT
TGAAAAGGGCCTTGCTAGCCTAAAAGGGGAAAAGGGCCTTGAAAAGCTCCTAAAAGGGAGGGCCTTGAAAAGG
GCCTTGAAAGAGCTAAAAGGGAGGGCCTTGAAAAGGGCCTTGAAAAGGGCCTAAGCCTAAAAGGGTAAAAGGGC
CTAAAAGGGCCTAAAAGGGCCTAAAAGGGCCTAAAAGGGCCTAAAAGGGGGGAAGGGCCTGGGAAGGGCCTGG
GATAGCCTAAAAGGGCCTAAAAGGGTTGGCAAGGGCCTAAAAGGGCCTAAAAGGGCCTAAAAGGGGGGAAGGG
CCTGGGAAGGGCCTGGGATAGCCTAAAAGGGGCCTAAAAGGGTTGGTTGAAAAGGGCCTTGAAAAGGGCCTTG
CCTAGCCTAAAAGGCCGGCCTAAAAGGGCCTAAAAGGGTATATATACACACACTAGAGAGAGAGATATATTATA
CACACACCATATATAAAAAATTTATATGGGTTCTCC○○○○○CCTAGCCTAAAAGGCCGGCCTAAAAGGGCCTAAA
GGGAAA



TABLE 15.2 SOME EUKARYOTIC GENOMES

Organism		Estimated Genome Size (Mbp)	Number of Genes (× 1,000)	Nature of Genome
<i>Vertebrates</i>				
	<i>Homo sapiens</i> (human)	3,200	20–25	The first large genome to be sequenced; the number of transcribable genes is far less than expected; much of the genome is occupied by repeated DNA sequences.
	<i>Pan troglodytes</i> (chimpanzee)	2,800	20–25	There are few base substitutions between chimp and human genomes, less than 2%, but many small sequences of DNA have been lost as the two species diverged, often with significant functional consequences.
	<i>Mus musculus</i> (mouse)	2,500	25	Roughly 80% of mouse genes have a functional equivalent in the human genome; importantly, large portions of the noncoding DNA of mouse and human have been conserved; overall, rodent genomes (mouse and rat) appear to be evolving more than twice as fast as primate genomes (humans and chimpanzees).
	<i>Gallus gallus</i> (chicken)	1,000	20–23	One-third the size of the human genome; genetic variation in domestic chickens seems much higher than in humans.
	<i>Fugu rubripes</i> (pufferfish)	365	35	The <i>Fugu</i> genome is only one-ninth the size of the human genome, yet it contains 10,000 more genes.
<i>Invertebrates</i>				
	<i>Caenorhabditis elegans</i> (nematode)	97	21	The fact that every cell of <i>C. elegans</i> has been identified makes its genome a particularly powerful tool in developmental biology.
	<i>Drosophila melanogaster</i> (fruit fly)	137	13	<i>Drosophila</i> telomere regions lack the simple repeated segments that are characteristic of most eukaryotic telomeres. About one-third of the genome consists of gene-poor centric heterochromatin.

Genome Sequence Similarities to Other Species

Fruit Fly
44%



Mouse
92%



Yeast
26%



Plant
18%

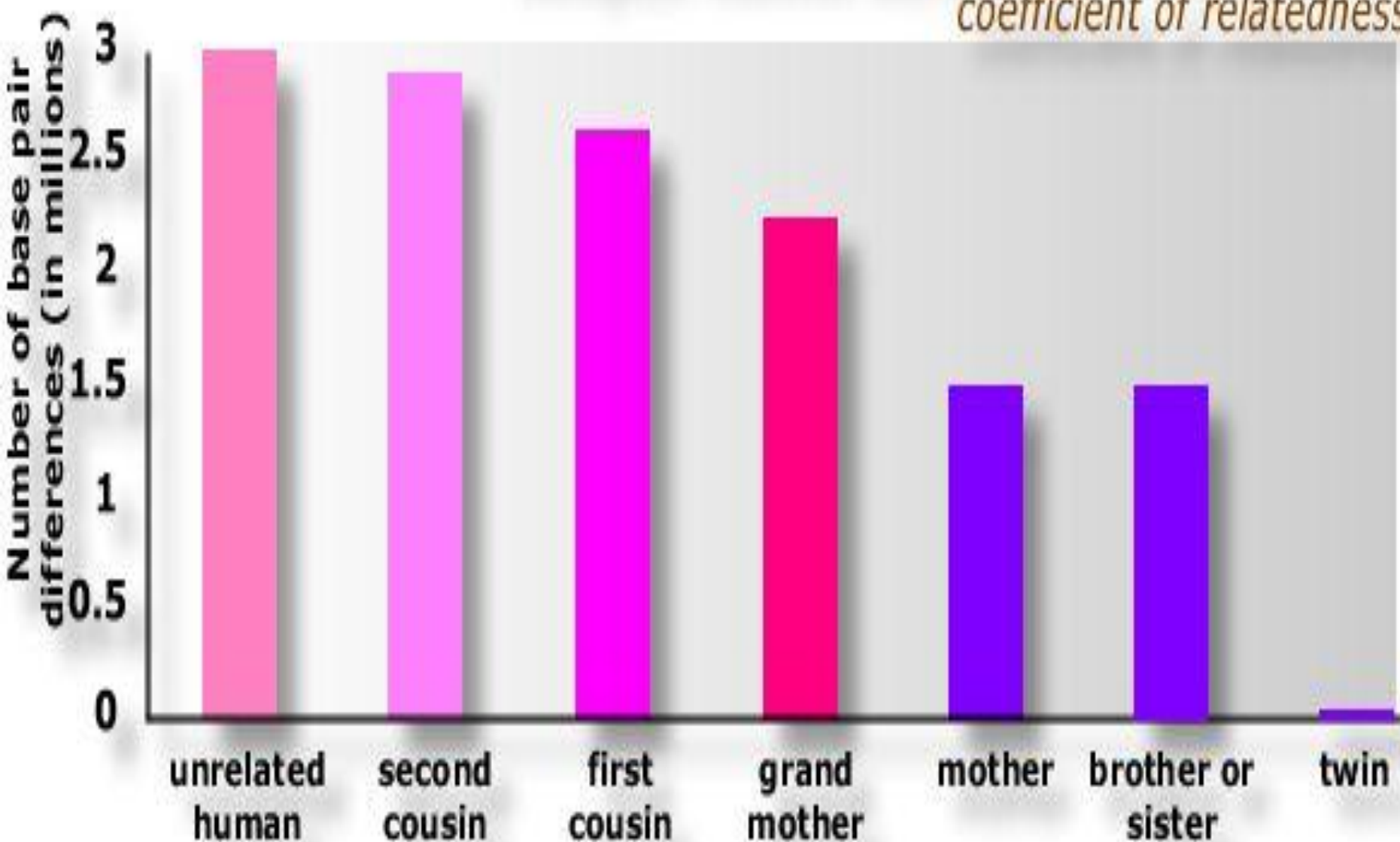


Chimp
98%



**What percent
of your genes
do you share?**

The number of base pair differences between biological relatives decreases as a function of their coefficient of relatedness.



Quantitative Advances Since the Human Genome Project (HGP)

	HGP Begins	HGP Ends	10 Years after HGP
			
Genome Sequencing			
Cost to Generate a Human Genome Sequence	~\$1 billion	~\$10-50 million	~\$3-5 thousand
Time to Generate a Human Genome Sequence	~6-8 years	~3-4 months	~1-2 days
Human Genome Sequences	0	1	Thousands

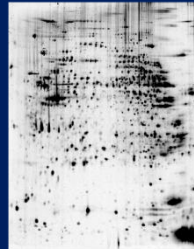
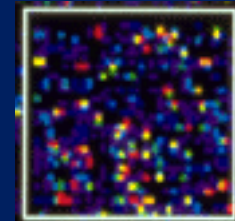
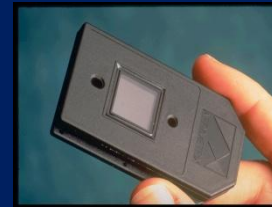
Quantitative Advances Since the Human Genome Project (HGP)

Genome Sequence Data			
Total DNA Bases in GenBank	~49 million	~31 terabases	~150 terabases
Whole-Genome Shotgun Bases in GenBank	0	~9.6 terabases	~391 terabases
Vertebrate Genome Sequences	0	3	112
Non-Vertebrate, Eukaryotic Genome Sequence	0	14	455
Prokaryotic Genome Sequences	0	167	8760
Human Single-Nucleotide Polymorphisms	~4.4 thousand	~3.4 million	~53.6 million

Genomic Information

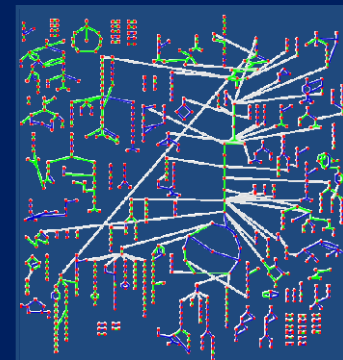
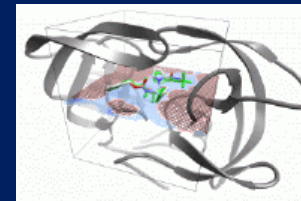
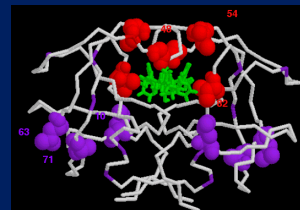
Novel Diagnostics

Microchips & Microarrays - DNA
Gene Expression - RNA
Proteomics - Protein



Novel Therapeutics

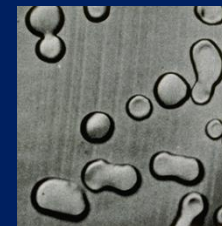
Drug Target Discovery
Rational Drug Design
Molecular Docking
Gene Therapy
Stem Cell Therapy



Understanding Metabolism

Understanding Disease

Inherited Diseases - OMIM
Infectious Diseases
Pathogenic Bacteria
Viruses



سر فصل ها

- تعریف و ارایه نتایج پروژه ژنوم انسان
- منابع اطلاعات مربوط به پروژه ژنوم انسان
- بیوانفورماتیک و آنالیز اطلاعات مربوط به پروژه ژنوم انسان
- تشخیص زودرس سرطان تخمدان با استفاده از بیوانفورماتیک

National Center for Biotechnology Information (NCBI)

[Genomes & Maps](#)

[Data & Software](#)

[DNA & RNA](#)

[Domains & Structures](#)

[Genes & Expression](#)

[Genetics & Medicine](#)

[Homology](#)

[Literature](#)

[Proteins](#)

[Sequence Analysis](#)

[Chemicals & Bioassays](#)

[Taxonomy](#)

[Training & Tutorials](#)

[Variation](#)

Genome Browsers - International

United States:

- <http://www.ncbi.nlm.nih.gov/mapview/> NIH-NCBI
- <http://genome.ucsc.edu/> University of California Santa Cruz
- [Genome.wustl.edu](http://genome.wustl.edu/) / Washington University

Europe:

- <http://www.ensembl.org/index.html> (The European Bioinformatics Institute). Part of the European Molecular Biology Laboratory) AND Wellcome trust Sanger Institute)

Japan:

- <http://utgenome.org/UTGBMedaka> University of Tokyo Genome Browser



BLAST®

Basic Local Alignment Search Tool

My NCBI



Welcome brutlag. [\[Sign Out\]](#)

[Home](#)

[Recent Results](#)

[Saved Strategies](#)

[Help](#)

► [NCBI/ BLAST Home](#)

BLAST finds regions of similarity between biological sequences. [more...](#)

New Aligning Multiple Protein Sequences? Try the [COBALT Multiple Alignment Tool](#). [Go](#)

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- | | | |
|---|--|--|
| ☐ Human | ☐ Oryza sativa | ☐ Gallus gallus |
| ☐ Mouse | ☐ Bos taurus | ☐ Pan troglodytes |
| ☐ Rat | ☐ Danio rerio | ☐ Microbes |
| ☐ Arabidopsis thaliana | ☐ Drosophila melanogaster | ☐ Apis mellifera |

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#)

Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontinuous megablast

[protein blast](#)

Search **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast

[blastx](#)

Search **protein** database using a **translated nucleotide** query

[tblastn](#)

Search **translated nucleotide** database using a **protein** query

[tblastx](#)

Search **translated nucleotide** database using a **translated nucleotide** query

News

SOAP BLAST

A SOAP based BLAST service is available.

Mon, 18 Jul 2011 08:00:00 EST

[More BLAST news...](#)

Tip of the Day

[Use Genomic BLAST to see the genomic context](#)

If you are interested in the evolution of a particular gene or gene family it is often interesting to examine the intro-exon structure even across species.

[More tips...](#)

سر فصل ها

- تعریف و ارایه نتایج پروژه ژنوم انسان
- منابع اطلاعات مربوط به پروژه ژنوم انسان
- بیوانفورماتیک و آنالیز اطلاعات مربوط به پروژه ژنوم انسان
- تشخیص زودرس سرطان تخمدان با استفاده از بیوانفورماتیک

Pre-genomic bioinformatics

"The mathematical, statistical and computing methods that aim to solve biological problems using DNA and Amino Acid sequences and related information."

According to this scheme, the monomers in a given macromolecule of DNA or protein can be treated computationally as *letters of an alphabet*, put together in pre-programmed arrangements to carry messages or do work in a cell.

Post-genomic bioinformatics

- Possess multiple whole genomes
- Look for differences and similarities between all the genes of multiple species.
- Draw particular conclusions about species and general ones about evolution.
- Referred to as *comparative genomics*

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

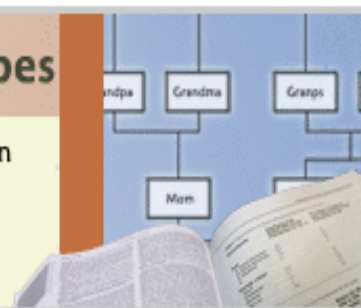
[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [NCBI News](#)

Get Started

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

Genotypes and Phenotypes

Data from Genome Wide Association studies that link genes and diseases. See study variables, protocols, and analysis.



FileEditViewHistoryBookmarksToolsHelp

Genomes & Maps - Site Guide - NCBI

www.ncbi.nlm.nih.gov/guide/genomes-maps/

33°C

Get Media Player Plugin

YouTubeNEWS

Facebook

NCBI

Resources

How To

NCBI

National Center for
Biotechnology Information

All Databases

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Genomes & Maps

AllDatabasesDownloadsSubmissionsToolsHow To

Databases

[BioProject \(formerly Genome Project\)](#)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

[CloneDB \(formerly Clone Registry\)](#)

A database that integrates information about clones and libraries, including sequence data, map positions and distributor information.

[Database of Genome Survey Sequences \(dbGSS\)](#)

A division of GenBank that contains short single-pass reads of genomic DNA. dbGSS can be searched directly through the Nucleotide GSS Database.

[Database of Genomic Structural Variation \(dbVar\)](#)

The dbVar database has been developed to archive information associated with large scale genomic variation, including large insertions, deletions, translocations and inversions. In addition to archiving variation discovery, dbVar also stores associations of defined variants with phenotype information.

[Epigenomics](#)

This resource enables users to explore and visualize richly-annotated epigenomics datasets. It provides a unique

Genomes & Maps - Si...Search Results1 - Paint

EN1:43 PM

Genomes & Maps - Site Guide - NCBI

www.ncbi.nlm.nih.gov/guide/genomes-maps/

Search The Web (privitize)

33°C

Get Media Player Plugin

YouTube

NEWS

Downloads

[FTP: Genome](#)

This site contains genome sequence and mapping data for organisms in Entrez Genome. The data are organized in directories for single species or groups of species. Mapping data are collected in the directory MapView and are organized by species. See the README file in the root directory and the README files in the species subdirectories for detailed information.

[FTP: Genome Mapping Data](#)

Contains directories for each genome that include available mapping data for current and previous builds of that genome.

[FTP: Genome Markers \(UniSTS\)](#)

This directory contains text and XML files for UniSTS records along with mapping data.

[FTP: RefSeq](#)

This site contains all nucleotide and protein sequence records in the Reference Sequence (RefSeq) collection. The ""release"" directory contains the most current release of the complete collection, while data for selected organisms (such as human, mouse and rat) are available in separate directories. Data are available in FASTA and flat file formats. See the README file for details.

[FTP: SKY/M-Fish and CGH Data](#)

This site contains SKY-CGH data in ASN.1, XML and EasySKYCGH formats. See the skycghreadme.txt file for more information.

[FTP: Sequence Read Archive \(SRA\) Download Facility](#)

This site contains next-generation sequencing data organized by the submitted sequencing project.

[FTP: Trace Archive](#)

This site contains the trace chromatogram data organized by species. Data include chromatogram, quality scores, FASTA sequences from automatic base calls, and other ancillary information in tab-delimited text as well as XML formats. See the README file for details.

Start

Microsoft PowerPoint - [...]

Genomes & Maps - Si...

Search Results

2 - Paint

through the Nucle

1:44 PM

Index of ftp://ftp.ncbi.nlm.nih.gov/genomes/

📁 Up to higher level directory

Name	Size	Last Modified
📁 ASSEMBLY_REPORTS		8/17/2013 8:31:00 AM
📁 Acyrthosiphon_pisum		3/15/2012 12:00:00 AM
📁 Aedes_aegypti		9/21/2010 12:00:00 AM
📁 Ailuropoda_melanoleuca		9/6/2012 12:00:00 AM
📁 Amphimedon_queenslandica		3/15/2012 12:00:00 AM
📁 Anas_platyrhynchos		7/3/2013 10:23:00 PM
📁 Anolis_carolinensis		3/15/2012 12:00:00 AM
📁 Anopheles_gambiae		1/14/2008 12:00:00 AM
📁 Apis_florea		4/13/2012 12:00:00 AM
📁 Apis_mellifera		12/20/2012 12:00:00 AM
📁 Aplysia_californica		7/24/2013 10:41:00 PM
📁 Arabidopsis_lyrata		5/11/2011 12:00:00 AM
📁 Arabidopsis_thaliana		9/30/2008 12:00:00 AM
📁 BACENDS		5/13/2010 12:00:00 AM
📁 Bacteria		8/17/2013 10:19:00 AM
📁 Bacteria_DRAFT		4/29/2013 8:30:00 PM
📁 Bombus_impatiens		3/15/2012 12:00:00 AM
📁 Bombus_terrestris		3/15/2012 12:00:00 AM
📁 Bombyx_mori		7/3/2013 10:05:00 PM
📁 Bos_taurus		8/8/2013 7:01:00 PM

 Gallus_gallus	6/28/2013	11:30:00 PM
 Glycine_max	3/15/2012	12:00:00 AM
 Gorilla_gorilla	1/2/2013	12:00:00 AM
 HUMAN_MICROBIOM	4/19/2012	12:00:00 AM
 H_sapiens	8/16/2013	4:52:00 PM
 Heterocephalus_glaber	6/21/2013	8:02:00 PM
<i>Homo_sapiens</i>	12/14/2012	12:00:00 AM
 Hydra_magnipapillata	3/8/2013	3:44:00 AM
 IDS	8/16/2013	2:43:00 PM
 INFLUENZA	8/17/2013	9:11:00 AM
 Jaculus_jaculus	6/6/2013	4:56:00 PM
 Loxodonta_africana	3/15/2012	12:00:00 AM
 MITOCHONDRIA	12/16/2011	12:00:00 AM
 M_musculus	4/15/2013	7:00:00 PM
 Macaca_mulatta	9/7/2012	12:00:00 AM

View/download features around an object or...

www.ncbi.nlm.nih.gov/guide/howto/view-features-object/

Search The Web (privitize)

4m

Search»

33°C

Get Media Player Plugin

YouTube

NEWS

NCBI

Resources

How To

Sign in to NCBI

NCBI

National Center for Biotechnology Information

All Databases

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

How to: View/download features around an object or between two objects on a chromosome

Starting with...

A MARKER NAME, GENE NAME/SYMBOL, OR SEQUENCE ACCESSION NUMBER (e.g., NM_138635)

1. Begin on the [Map Viewer](#) home page. If necessary, expand the Search window by clicking the arrow, then select an organism and enter the search term. To display the region between two objects, use "OR" as follows: D10S275 OR D2S2560.

2. Click on the Map Element or a single Map for the desired assembly.

3. If necessary, use the [Maps & Options](#) dialog box to change displayed maps; the maps and region displayed determine the data available. You can also use the two boxes under Region Shown in the blue sidebar on the left to enter names, accessions, and coordinates.

4. Click [Data as Table View](#) in the blue sidebar on the left.

5. Each map has a link to Download Data as a tab-delimited text file.

CHROMOSOMAL COORDINATES

1. Begin on the [Map Viewer](#) home page. Click the "R" icon under Tools for the desired organism and build.

2. Select the chromosome, enter the coordinates in the From and To boxes, and click Go. Use either exact coordinates, e.g., 61551076, or values such as, 61M or 61551K.

3. If necessary, use the [Maps & Options dialog box](#) to change displayed maps; the maps and region displayed determine the data available.

4. Click [Data as Table View](#) in the blue sidebar on the left.

5. Each map has a link to Download Data as a tab-delimited text file.

Start

Microsoft PowerPoint - [...]

Index of ftp://ftp.ncbi.nlm...

View/download featu...

Search Results

7 - Paint

WinZip - hs_alt_CHM1_1...

EN

2:07 PM

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More..](#)

Search

Search:

for:

Tools Legend

- Search or Browse the Genome
- BLAST
- Clone Finder
- Go to region on a chromosome
- Genome Resources page

News

- Mouse annotation release 103** Apr 3
Mouse annotation release 103 has been released to Mapviewer... [more](#)
- Ovis aries annotation release 100** Jan 29
Ovis aries annotation release 100 has been released to mapvi... [more](#)
- Rattus norvegicus build 5.1 released** Jan 9
Rattus norvegicus build 5.1 has been released to mapviewer... [more](#)
- Human annotation release 104** Dec 3, 2012
Human annotation release 104 released to mapviewer. The chro... [more](#)

[Show all](#)

Vertebrates (36)

Mammals (27)

Primates (11)

Scientific name	Common name	Build	Tools
<i>Callithrix jacchus</i>	white-tufted-ear marmoset	Annotation Release 101	
<i>Gorilla gorilla</i>	western gorilla	Annotation Release 100	
<i>Homo sapiens</i>	human	Annotation Release 104 Build 36.3	
<i>Macaca mulatta</i>	rhesus macaque	Build 1.2	
<i>Nomascus leucogenys</i>	northern white-cheeked gibbon	Annotation Release 101 Build 1.1	
<i>Otlemur garnettii</i>	small-eared galago	Annotation Release 100	
<i>Pan paniscus</i>	pygmy chimpanzee	Annotation Release 100	
<i>Pan troglodytes</i>	chimpanzee	Annotation Release 102 Build 2.1	
<i>Papio anubis</i>	olive baboon	Annotation Release 100	
<i>Pongo abelii</i>	Sumatran orangutan	Annotation Release 101	
<i>Saimiri boliviensis</i>	Bolivian squirrel monkey	Annotation Release 100	

Rodents (4)

Scientific name	Common name	Build	Tools
<i>Cavia porcellus</i>	domestic guinea pig	Build 1.1	
<i>Cricetus griseus</i>	Chinese hamster	Build 1.1	
<i>Mus musculus</i>	laboratory mouse	Annotation Release 103 Build 37.2	
<i>Rattus norvegicus</i>	rat	Build 5.1 (Annotation Release 103) RGSC v3.4	

Monotremes (1)

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More..](#)

Search ▼

Search:

for:

Go

- Search Legend ▼
- Search or Browse the Genome
 - BLAST
 - Clone Finder
 - Go to region on a chromosome
 - Genome Resources page

- News ▼
- Genome annotation release 103** Apr 3
Genome annotation release 103 has been released to Mapviewer.... [more](#)
 - Genomes annotation release 100** Jan 29
Genomes annotation release 100 has been released to mapvi... [more](#)
 - Genus norvegicus build 5.1 released** Jan 9
Genus norvegicus build 5.1 has been released to viewer... [more](#)

▼ **Vertebrates**

▼ **Mammals**

▼ **Primates**

Scientific name	Common name	Build
<i>Callithrix jacchus</i>	white-tufted-ear marmoset	Annotation Release 101
<i>Gorilla gorilla</i>	western gorilla	Annotation Release 100
<i>Homo sapiens</i>	human	Annotation Release 104 Build 36.3
<i>Macaca mulatta</i>	rhesus macaque	Build 1.2
<i>Nomascus leucogenys</i>	northern white-cheeked gibbon	Annotation Release 101 Build 1.1
<i>Otolemur garnettii</i>	small-eared galago	Annotation Release 100
<i>Pan paniscus</i>	pygmy chimpanzee	Annotation Release 100
<i>Pan troglodytes</i>	chimpanzee	Annotation Release 102 Build 2.1
<i>Papio anubis</i>	olive baboon	Annotation Release 100
<i>Pongo abelii</i>	Sumatran orangutan	Annotation Release 101
<i>Saimiri boliviensis</i>	Bolivian squirrel monkey	Annotation Release 100

▼ **Rodents**

Scientific name	Common name	Build
<i>Cavia porcellus</i>	domestic guinea pig	Build 1.1
<i>Cricetulus griseus</i>	Chinese hamster	Build 1.1
<i>Mus musculus</i>	laboratory mouse	Annotation Release 103 Build 37.2

Map Viewer

Map Viewer Home
Map Viewer Help
Human Maps Help
Release Notes

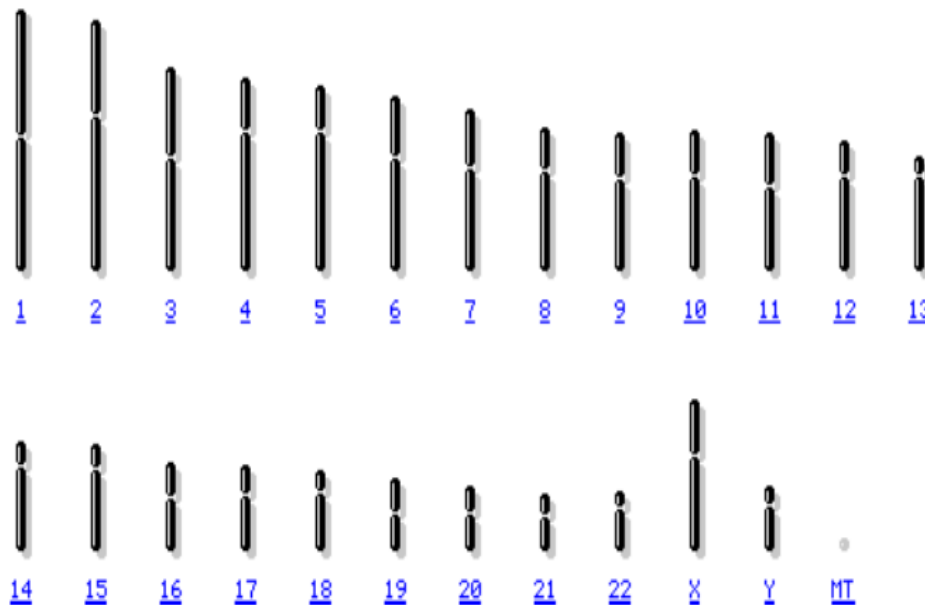
NCBI Resources

Genome Project
Trace Archive
(Venter)
RefSeq
Whole Genome
Association (WGA)
Human Genome
Resources
GRC
Consensus CoDing
Sequence (CCDS)

Homo sapiens (human) genome view

[BLAST search the h](#)

[Annotation Release 104 statistics](#) [Switch to previous build](#)



Lineage: [Eukaryota](#); [Metazoa](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Euteleostomi](#); [Mammalia](#); [Eutheria](#); [Euarchon](#); [Primates](#); [Haplorrhini](#); [Catarrhini](#); [Hominidae](#); [Homo](#); [Homo sapiens](#)

November 2012, Human Annotation Release 104

[Homo sapiens \(human\)](#) [Annotation Release 104 \(Current\)](#)

[BLAST human](#)

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

Master Map: Genes On Sequence

[Summary of Maps](#)

[Download/View Sequ](#)

Region Displayed: 0-250Mbp

Ideogram	Contig	Regions	Hs Uni6	Genes_seq	Symbol	Q	Links	E	Cyto	Description
	NT_077402.2	REGION77	Hs_735988	TP73	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p36.3	tumor protein p73	
	NT_077912.1		Hs_732228	MTOR	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p36.2	mechanistic target of rapamycin (serine/threonine kinase)	
	NT_044255..		Hs_097785	MTHFR	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p36.3	OTTHUMP0000002368	
	NT_021957..			NPPB	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p36.2	natriuretic peptide B	
	NT_044616..	REGION69	Hs_732467	HDAC1	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p34	histone deacetylase 1	
			Hs_678089	JUN	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p32-p31	jun proto-oncogene	
			Hs_732269	LEPR	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p31	leptin receptor	
	NT_052977..9		Hs_146542	F3	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p22-p21	coagulation factor III (thromboplastin, tissue factor)	
			Hs_748425	GSTM1	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p13.3	glutathione S-transferase mu 1	
			Hs_732392	PTPN22	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1p13.2	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	
			Hs_732646	MUC1	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q21	mucin 1, cell surface associated	
			Hs_035034	LMNA	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q22	lamin A/C	
	NT_077389.3	IP11	Hs_734495	CRP	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q23.2	C-reactive protein, pentraxin-related	
	NT_113793.2		Hs_732431	F5	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q23	coagulation factor V (proaccelerin, labile factor)	
	NT_113796.2		Hs_748689	FASLG	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q23	Fas ligand (TNF superfamily, member 6)	
	NT_113797.1		Hs_748676	PTGS2	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q25.2-q25.3	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase)	
	NT_079485.4	1Q21	Hs_732298	CFH	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q32	complement factor H	
	NT_079497.3	REGION2	Hs_505378	IL10	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q31-q32	interleukin 10	
	NT_079652.2	NT11	Hs_731785	PARP1	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q41-q42	poly (ADP-ribose) polymerase 1	
	NT_167185.1		Hs_733620	AGT	+	OMIM HGNC sv pr dlev hmsts CCDSSNP	best RefSeq	1q42.2	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	
	NT_113799.1		Hs_732047							
	NT_044487..		Hs_748357							
			Hs_737616							
	NT_044602.2	REGION3	Hs_655974							
		REGION70	Hs_748425							
		1Q32	Hs_732319							
	NT_167186.1	REGION79	Hs_732015							
			Hs_732499							
			Hs_189514							
			Hs_738168							
			Hs_567671							

Summary of Maps:

Map 1: Ideogram

JUN jun proto-oncogene [Homo sapiens (human)]

Gene ID: 3725, updated on 11-Aug-2013

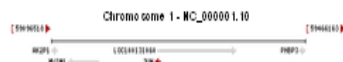
Summary

Official Symbol: JUN provided by HGNC
Official Full Name: Jun proto-oncogene provided by HGNC
Primary source: HGNC:5204
See related: Ensembl:ENSG00000177505; HPRD:01302; MIM:165160; Vega:0.TTTHUM000000008375
Gene type: protein coding
RefSeq status: REF/VEP
Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhina; Catarrhini; Hominoidea; Homo
Also known as: AP-1; AP-1; c-Jun
Summary: This gene is the putative transforming gene of rat sarcoma virus 17. It encodes a protein which is highly similar to the viral protein, and which interacts directly with specific target DNA sequences to regulate gene expression. This gene is linkless and is mapped to 1p32-p31, a chromosomal region involved in both translocations and deletions in human malignancies. (provided by RefSeq, Jul 2008)

Genomic context

Location: 1p32-p31
Sequence: Chromosome: 1; NC_000001.10 (59246463..59249785, complement)

See JUN in BioGenomics Map Viewer

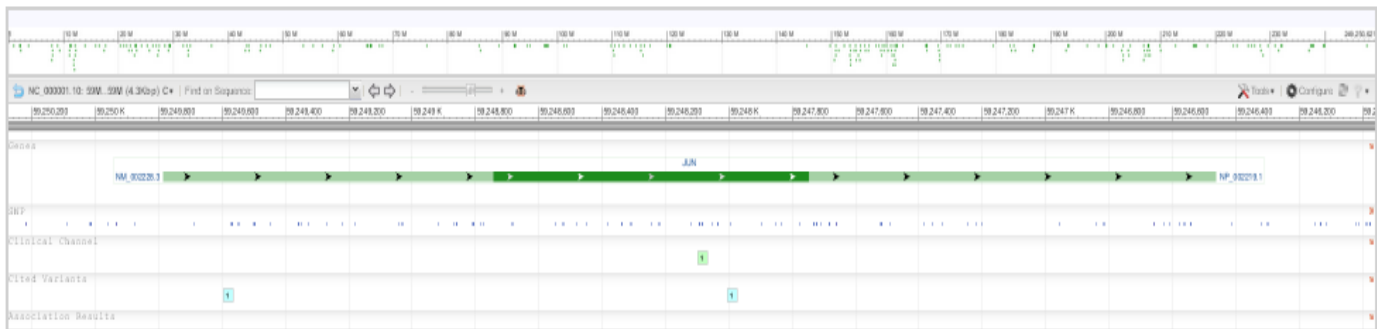


Genomic regions, transcripts, and products

Genomic Sequence NC_000001 chromosome 1 reference GRCh37.p10 Primary Assembly

Go to reference sequence details

Go to nucleotide Graphics FASTA GenBank



Bibliography

Related articles in PubMed

Table of contents

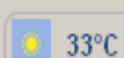
- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- RN-1 protein interactions
- Interactions
- Pathway
- General gene information
- Markers, Homology, Gene Ontology
- General protein information
- Reference sequences
- Related sequences
- Additional links

Related information

- Order cDNA clone
- 3D structures
- BioAssay
- BioAssay by Protein Target
- BioProtect
- BioSystems
- CCDC
- ClinVar
- Conserved Domains
- dbVar
- EST
- Full text in PMC
- Full text in PMC_nucleotide
- Gene neighbors
- Genome
- GED Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- Probe
- Protein
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed (Gene Ref)
- PubMed (OMIM)
- PubMed(nucleotide)(PMC)



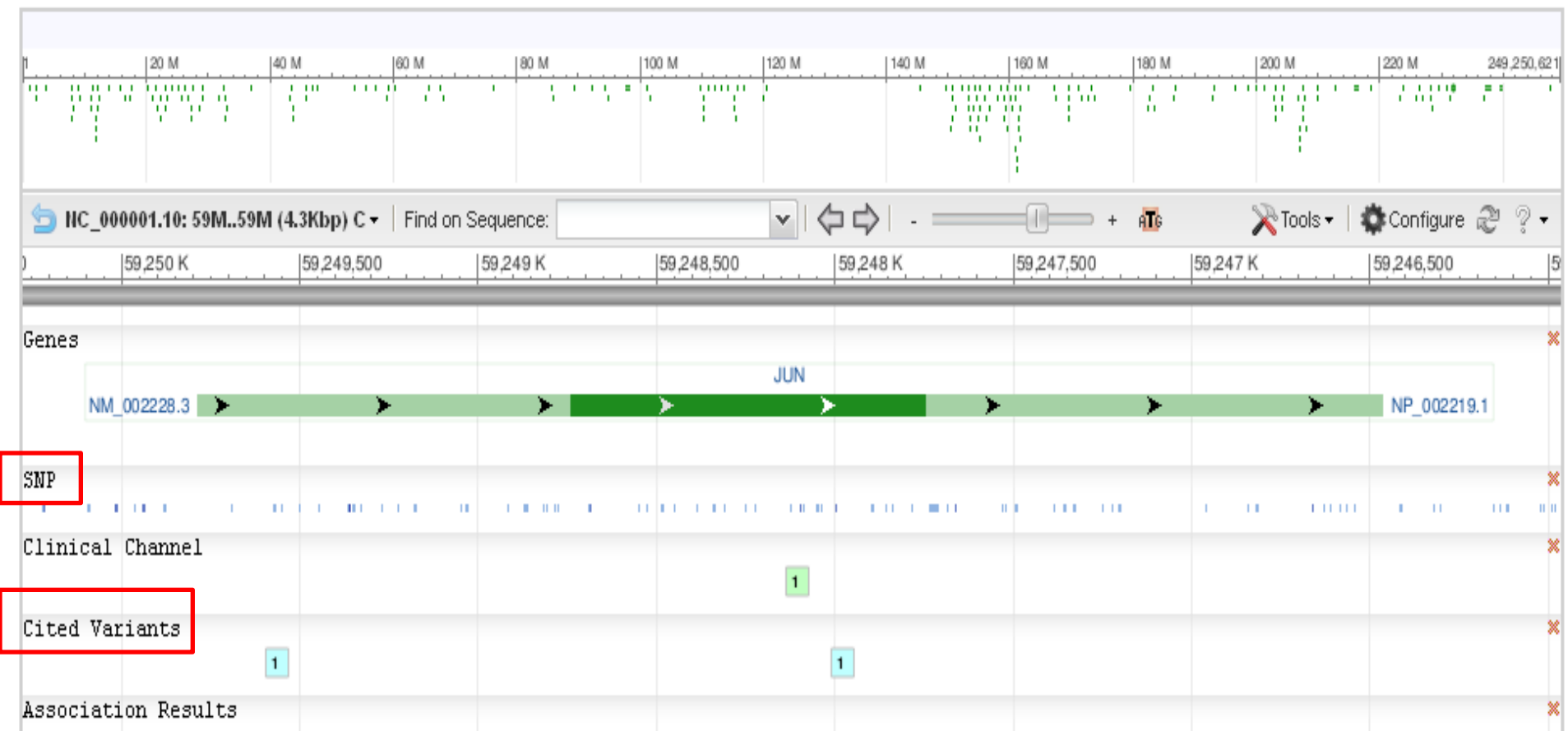
Search»



33°C

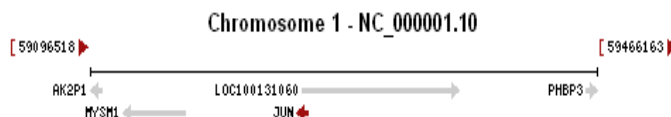
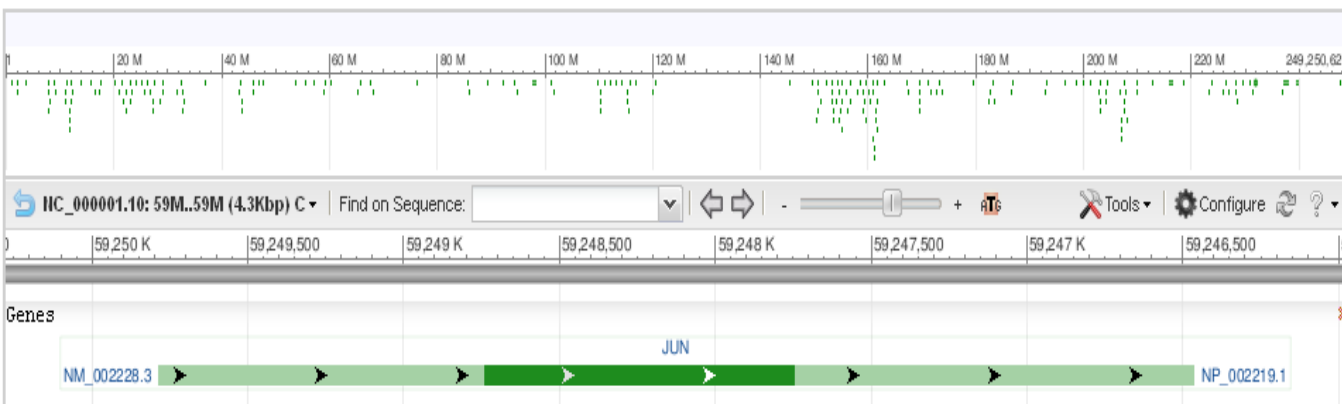
[Get Media Player Plugin](#)

Genomic regions, transcripts, and products

Genomic Sequence [Go to reference sequence details](#)[Go to nucleotide](#) [Graphics](#) [FASTA](#) [GenBank](#)

Bibliography



Also known as AP1; AP-1; c-Jun**Summary** This gene is the putative transforming gene of avian sarcoma virus 17. It encodes a protein which is highly similar to the viral protein, and which interacts directly with specific target DNA sequences to regulate gene expression. This gene is intronless and is mapped to 1p32-p31, a chromosomal region involved in both translocations and deletions in human malignancies. [provided by RefSeq, Jul 2008]**Genomic context****Location:** 1p32-p31**Sequence:** Chromosome: 1; NC_000001.10 (59246463..59249785, complement)See JUN in [Epigenomics](#), [MapViewer](#)**Genomic regions, transcripts, and products****Genomic Sequence** NC_000001 chromosome 1 reference GRCh37.p10 Primary AssemblyGo to [reference sequence details](#)Go to nucleotide [Graphics](#) [FASTA](#) [GenBank](#)**Additional links****Related information**[Order cDNA clone](#)[3D structures](#)[BioAssay](#)[BioAssay, by Protein Target](#)[BioProjects](#)[BioSystems](#)[CCDS](#)[ClinVar](#)[Conserved Domains](#)[dbVar](#)[EST](#)[Full text in PMC](#)[Full text in PMC_nucleotide](#)[Gene neighbors](#)[Genome](#)[GEO Profiles](#)[HomoloGene](#)[Map Viewer](#)[Nucleotide](#)[OMIM](#)[Probe](#)[Protein](#)

jun proto-oncogene [Homo sapiens (hu...]

www.ncbi.nlm.nih.gov/gene?cmd=retrieve&dopt=full_report&list_uids=3725

Search The Web (privitize)

Search»

33°C

Get Media Player Plugin

YouTube

NEWS

NM_002228.3

JUN

NP_002219.1

NP

linical Channel

ited Variants

ssociation Results

1

1

PubChem Substances

Probe

Protein

PubChem Compound

PubChem Substance

PubMed

PubMed (GeneRIF)

PubMed (OMIM)

PubMed(nucleotide/PMC)

RefSeq Proteins

RefSeq RNAs

SNP

SNP: GeneView

SNP: Genotype

Taxonomy

UniGene

UniSTS

Links to other resources

HGNC

Ensembl

HPRD

Vega

Evidence Viewer

AceView

PharmGKB

NCBI

Bibliography

Related articles in PubMed

1. Butyrate produced by commensal bacteria potentiates phorbol esters induced AP-1 response in human intestinal epithelial cells. Nepelska M, et al. PLoS One, 2012. PMID 23300800.

2. An AP-1 binding site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. Duverger A, et al. J Virol, 2013 Feb. PMID 23236059.

3. Human serum protein enhances HIV-1 replication and up-regulates the transcription factor AP-1. Perdomo MF, et al. Proc Natl Acad Sci U S A, 2012 Oct 23. PMID 23047699.

4. Aberrant expression of c-Jun in glioblastoma by internal ribosome entry site (IRES)-mediated translational activation. Blau L, et al. Proc Natl Acad Sci U S A, 2012 Oct 16. PMID 23027969.

5. Chitooligosaccharide elicits acute inflammatory cytokine response through AP-1 pathway in human intestinal epithelial-like (Caco-2) cells. Bahar B, et al. Mol Immunol, 2012 Jul. PMID 22512945.

See all (817) citations in PubMed

See citations in PubMed for homologs of this gene provided by HomoloGene

bi.nlm.nih.gov/pcsubstance?LinkName=gene_pcsubstance&from_uid=3725

Start

Index of ftp://ftp.ncbi.nlm.nih.gov/

JUN jun proto-oncogene...

Search Results

15 - Paint

Microsoft PowerPoint - [...]

2:37 PM

Genome Informatics

Chromosomes

Chromosome Abnormalities

Deoxyribonucleic Acid (DNA)

DNA Chip Technology

DNA Microarray Technology

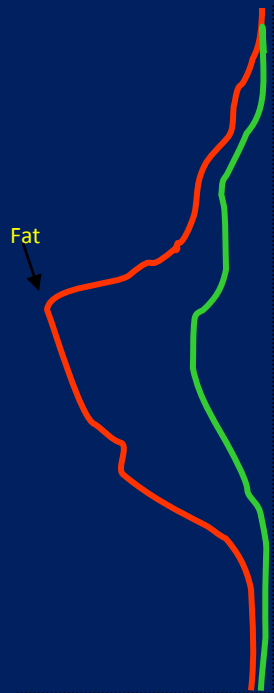
Epigenomics

Polymerase Chain Reaction (PCR)

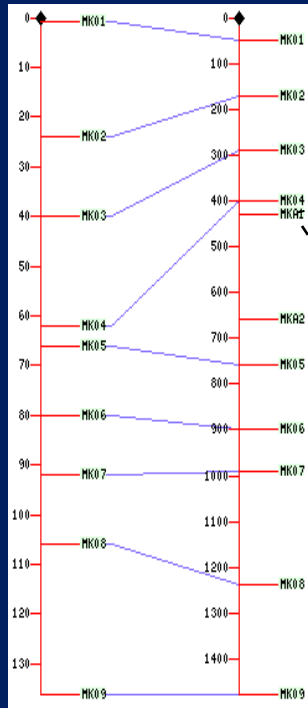
Transcriptome

Expression data & Supporting literature

Pig

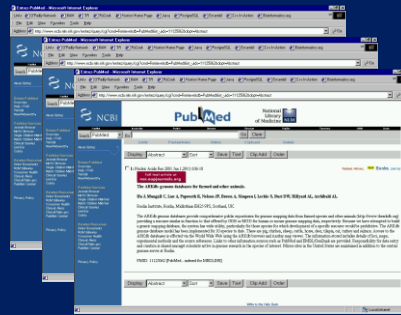


Trait location



Linkage Map

Radiation Hybrid Map

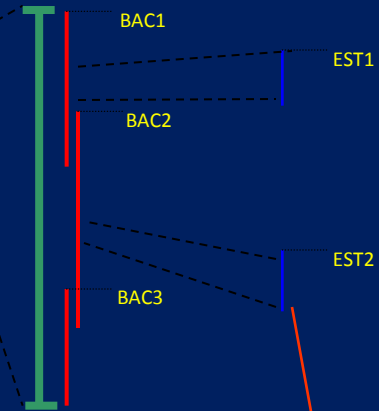


Linked References



Cytogenetic Map

Human



Physical Mapping

Chicken



Expression Analysis

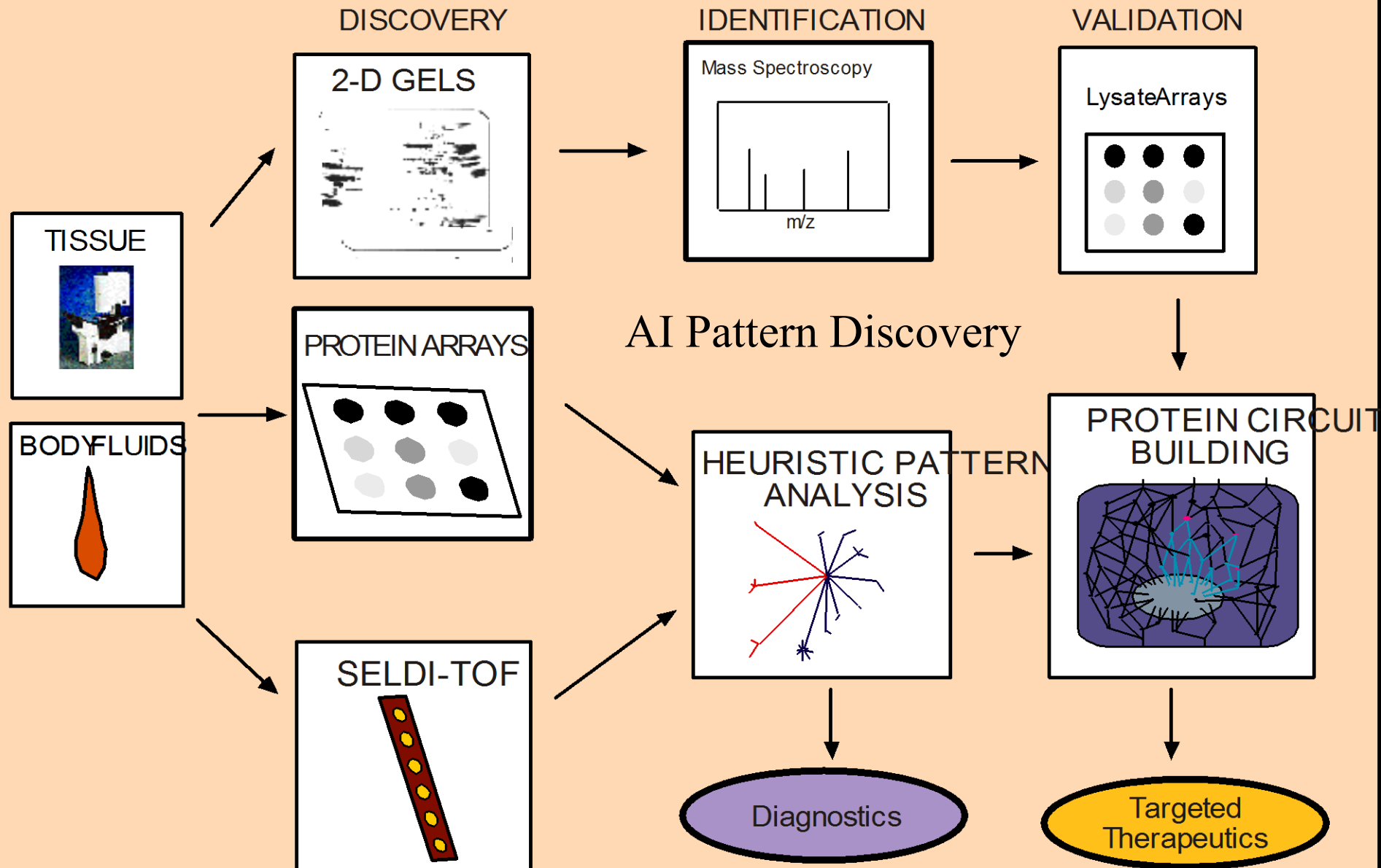
Analysis of the Data

- The *Genome sequence should be annotated* and the information should be structured so that *all sorts of queries can be run on the database.*
- * *Integrated and linked databases*
- * The *genetic variation database* should include or link to information on individual phenotypic variation.
- *Functional/expression database*, including *pathway/regulatory databases*

سر فصل ها

- تعریف و ارایه نتایج پروژه ژنوم انسان
- منابع اطلاعات مربوط به پروژه ژنوم انسان
- بیوانفورماتیک و آنالیز اطلاعات مربوط به پروژه ژنوم انسان
- تشخیص زودرس سرطان تخمدان با استفاده از بیوانفورماتیک

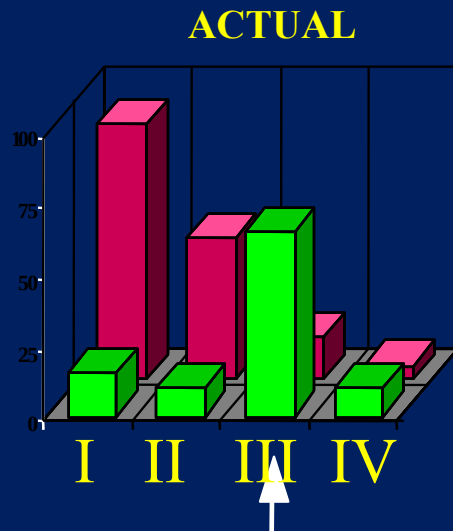
NCI / FDA CLINICAL PROTEOMICS PROGRAM



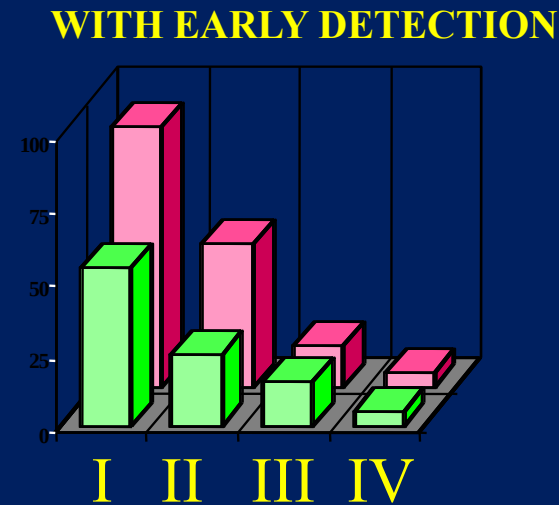
Proteomic information patterns in a drop of blood

Application to early detection of ovarian cancer

 Stage at diagnosis  5 Year Survival



Most ovarian cancers are detected after they have already spread beyond the ovary and the 5yr survival is less than 20%



Early diagnosis alone, with conventional therapy, can increase survival to greater than 95%

MECHANISMS OF DISEASE

Mechanisms of disease

Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

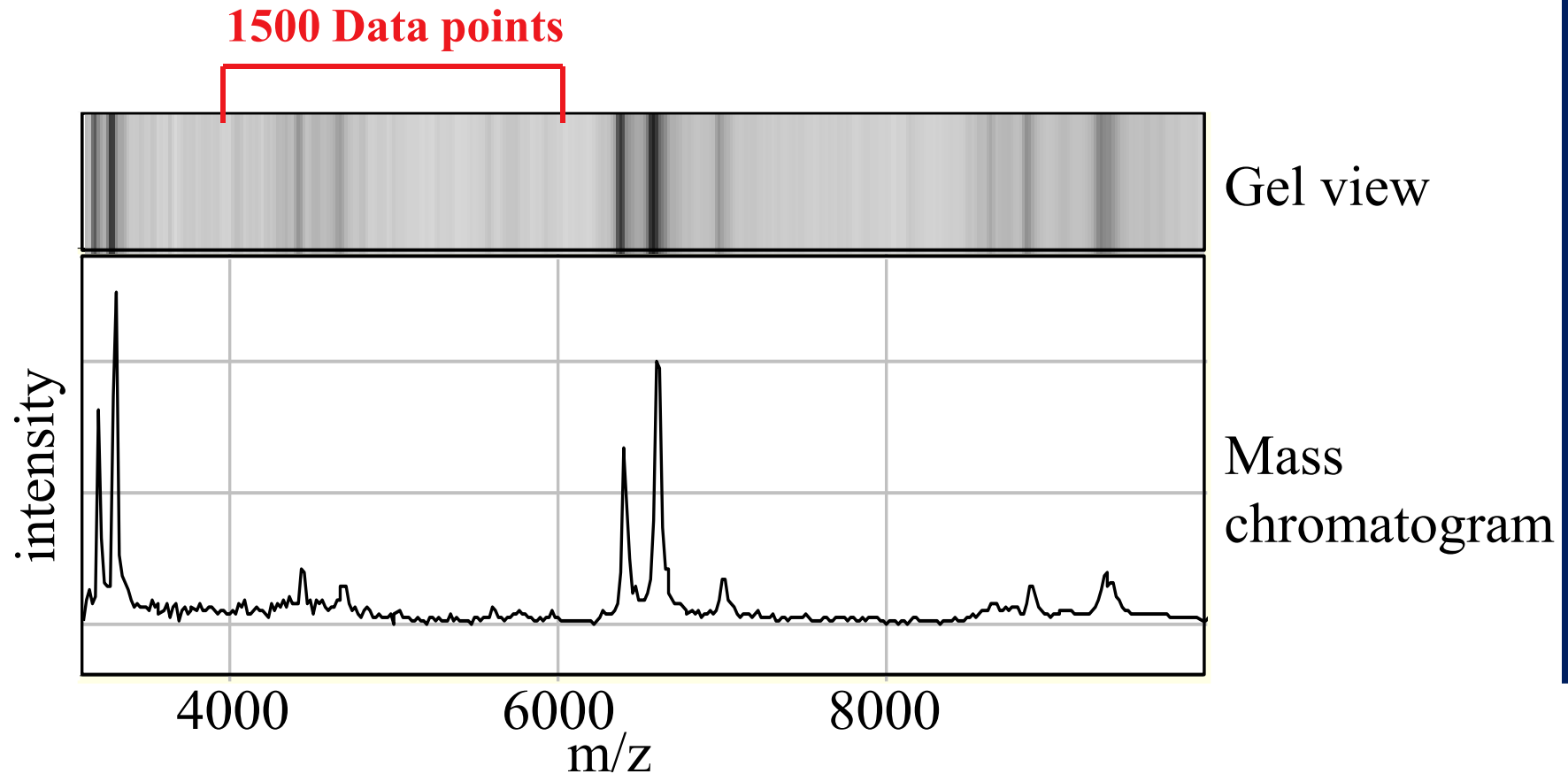
Summary

Background New technologies for the detection of early-stage ovarian cancer are urgently needed. Pathological changes within an organ might be reflected in proteomic patterns in serum. We developed a bioinformatics tool and used it to identify proteomic patterns in serum that distinguish neoplastic from non-neoplastic disease within the ovary.

Introduction

Application of new technologies for detection of ovarian cancer could have an important effect on public health,¹ but to achieve this goal, specific and sensitive molecular markers are essential.^{1–5} This need is especially urgent in women who have a high risk of ovarian cancer due to family or personal history of cancer, and for women with a genetic predisposition to cancer due to abnormalities

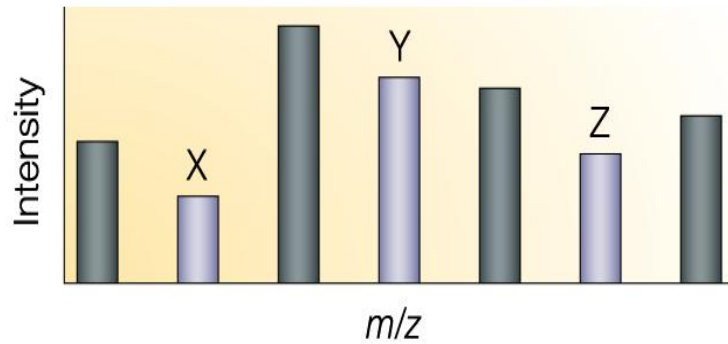
Typical serum profile from SELDI analysis



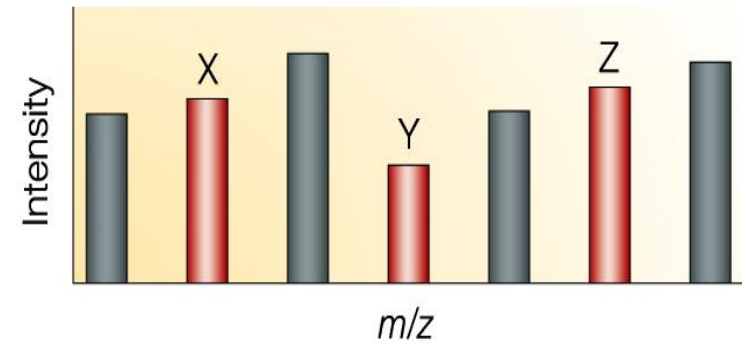
Data analysis window: 0-20,000 Daltons = 15,500 data points

Proteomic Pattern Clustering in n -Space

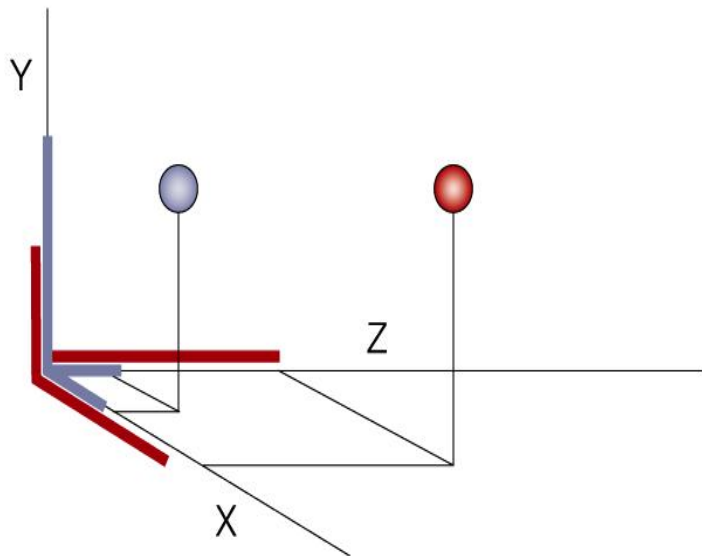
B Benign/unaffected spectra



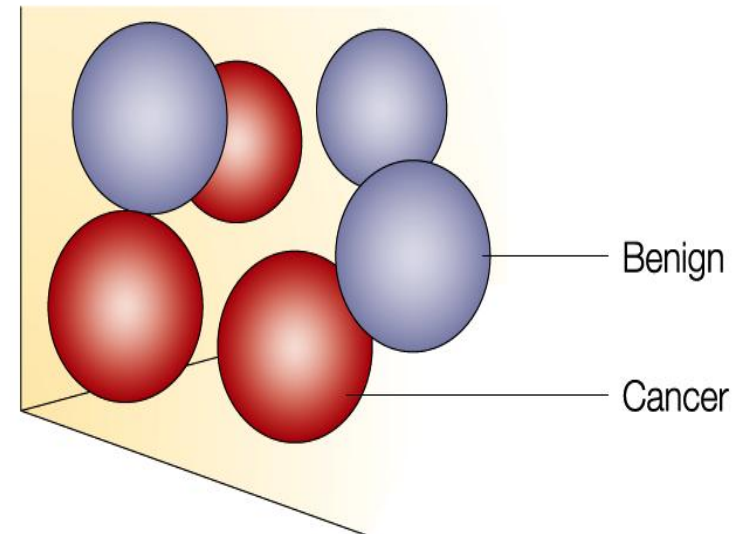
Cancer spectra



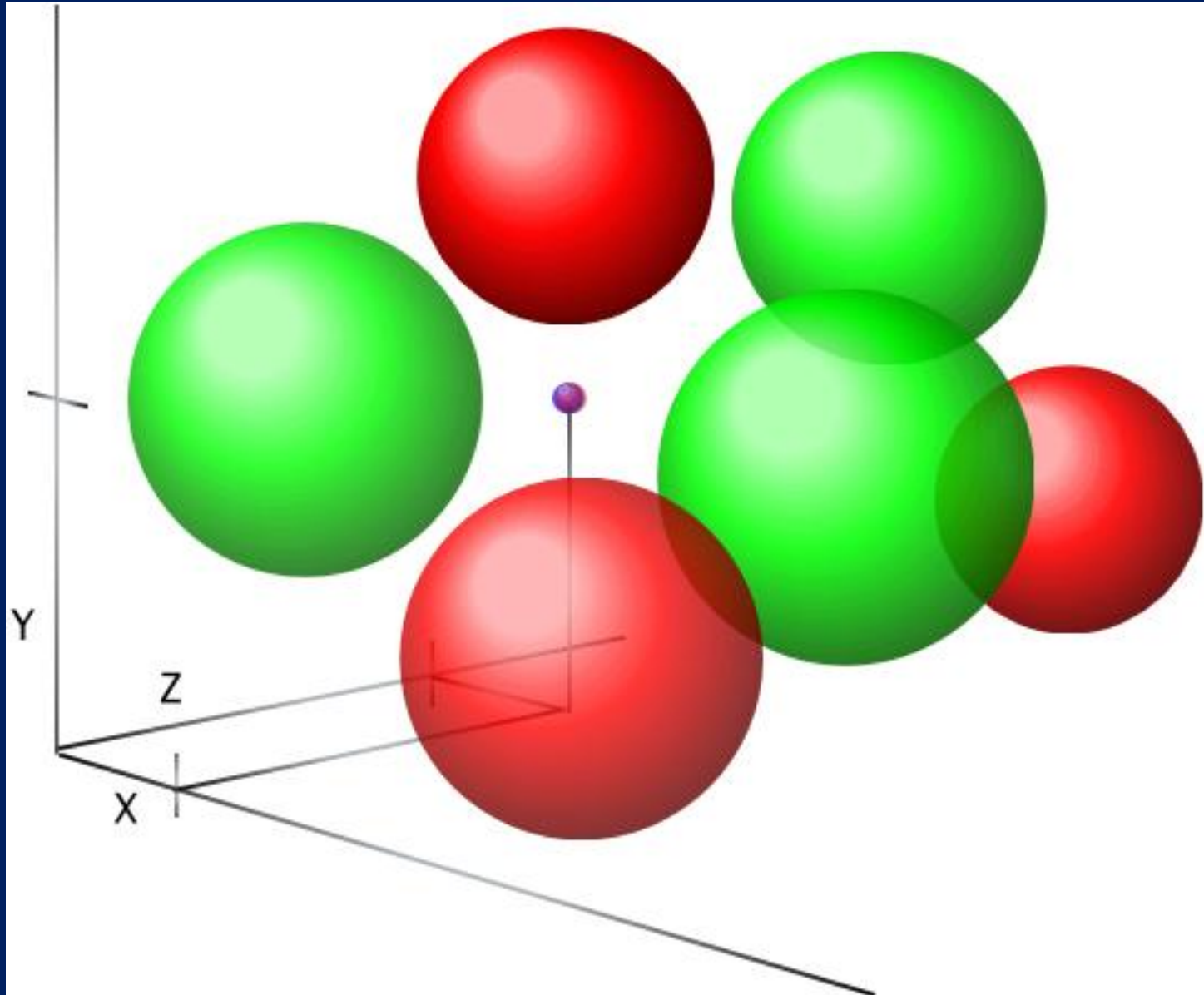
Plot of each pattern as a point in n -space



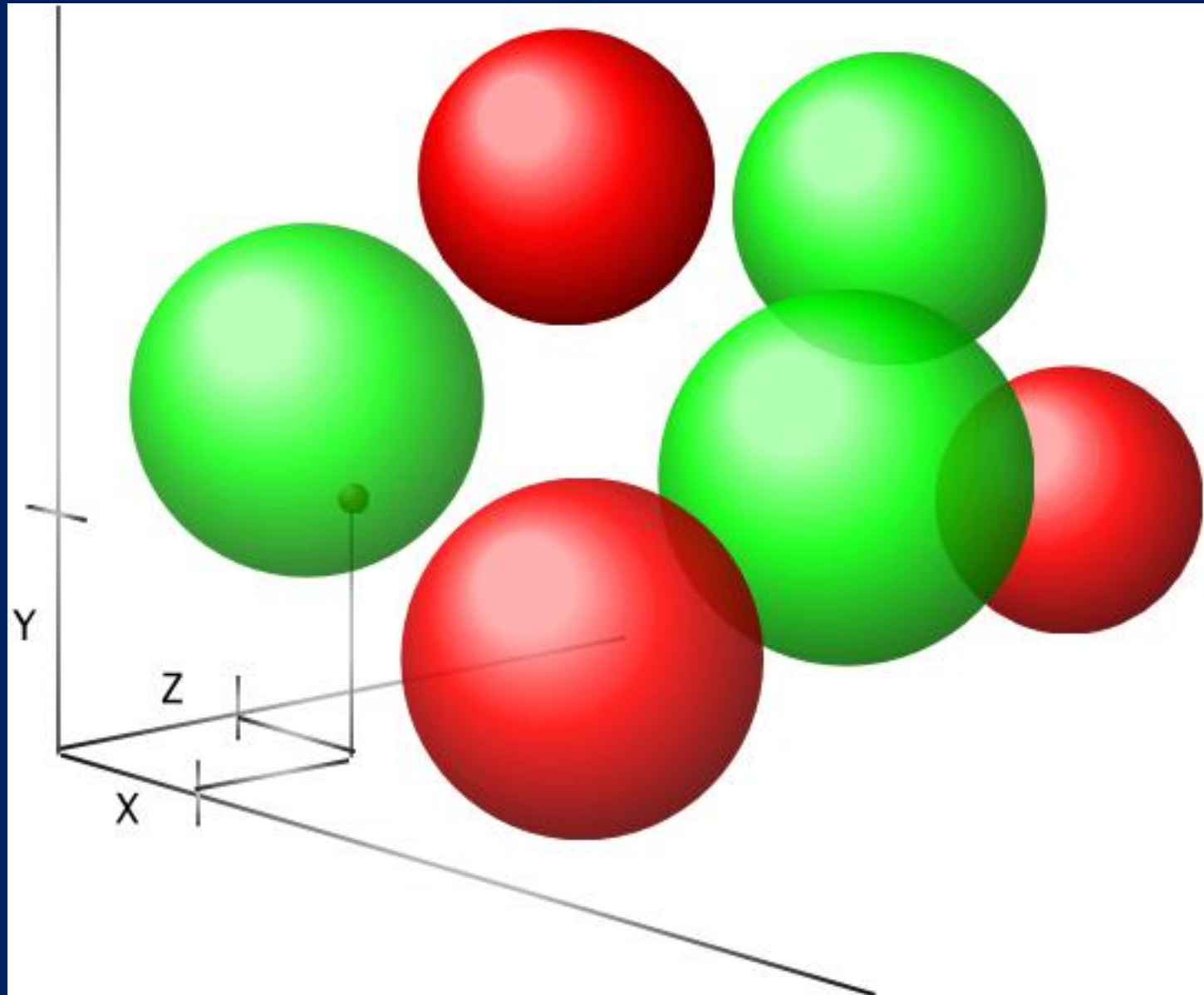
Clusters of points in n -space



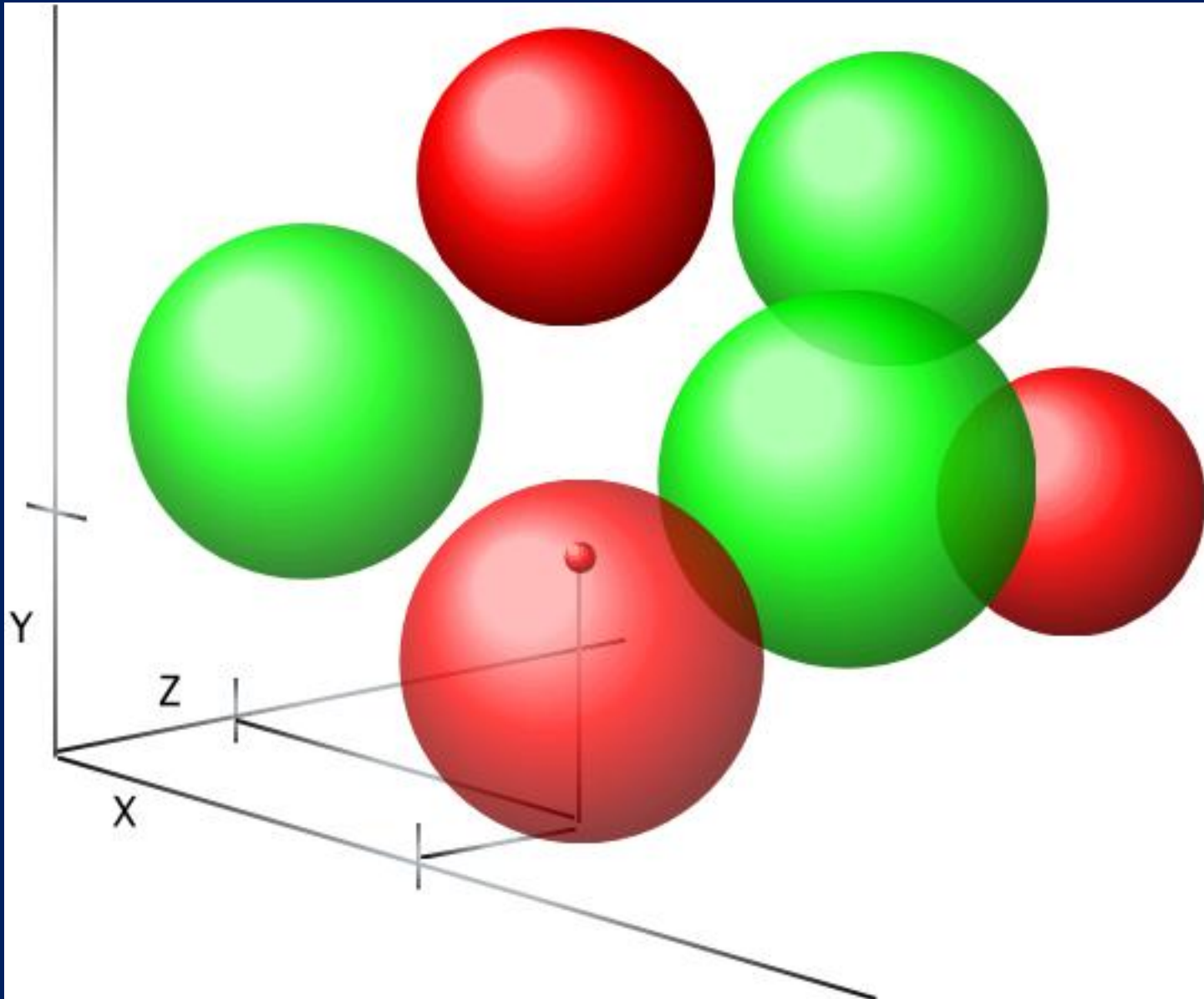
Classification of Unknown: NO MATCH



Classification of Unknown: BENIGN MATCH



Classification of Unknown: *CANCER MATCH*



Training Sets

OVARY UNAFFECTED TRAINING SET ^a	
DISEASE STATUS	N
No Evidence of Disease: No Evidence of Cysts	37
No Evidence of Disease: Benign Ovarian Cysts < 2.5 cm	11
No Evidence of Disease: Benign Ovarian Cysts > 2.5 cm	2
TOTAL	50

OVARY CANCER TRAINING SET ^b	
DISEASE STATUS	N
Surgically Staged Ovarian Cancer Stage II, III, IV	44
Surgically Staged Ovarian Cancer Stage I	6
TOTAL	50

a = Ovarian cysts diagnosed by 3-Dimensional ultrasound, no evidence of cancer during 5 year follow-up. Sera donated upon enrollment and prior to gynecologic evaluation.

b = Sera donated before surgical staging and diagnosis.

Artificial Intelligence-Based Proteomic Pattern Diagnostics

Ovarian Cancer Results

BLINDED TEST DATA (N= 116 PATIENTS)

NED (5 Year follow up) **22/24 (92%)**

Benign gynecologic and non-gyn inflammatory
(cysts, fibroids) (RA, colitis, sinusitis) **41/42 (98%)**

TOTAL SPECIFICITY: $63/66 = 95\%$

Ovarian cancer Stage II, III, IV **32/32 (100%)**

Ovarian cancer Stage I **18/18 (100%)**

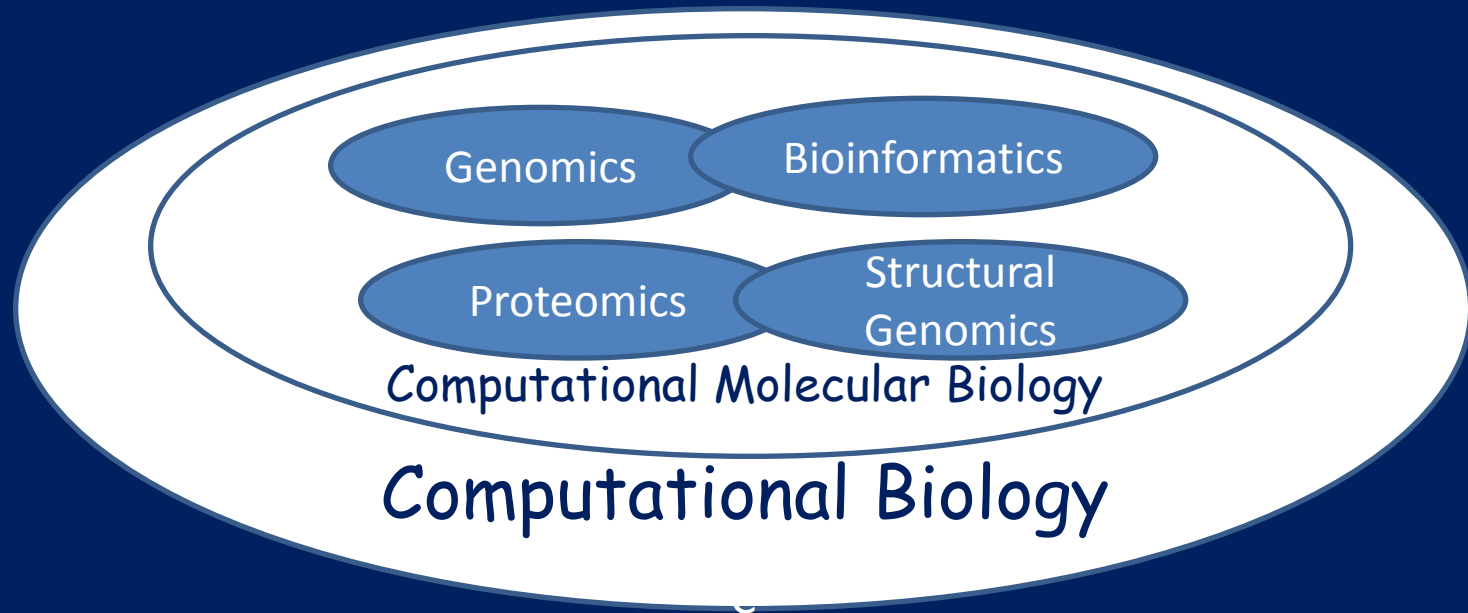
TOTAL SENSITIVITY: $50/50 = 100\%$

POSITIVE PREDICTIVE VALUE: 95% VS. 20% FOR CA125

The Promise of

**BIG
DATA**

Genomics, Bioinformatics & Computational Biology



Machine
learning

Robotics

Algorithms

Information
theory

Artificial
Intelligence

Databases

Graph
theory

Statistics &
probability