生物資訊簡介

Introduction to Bioinformatics

Hsuan-Ting Chang 張軒庭
EE Department, YunTech
國立雲林科技大學 電機系 光電與資訊實驗室

Outline

1. What is bioinformatics? 何謂生物資訊
2. Molecular genetics 分子生物學
3. Human genome project 人類基因體計畫
4. Databases & information retrieval 資料庫與資訊擷取
5. Global research issues 全球研究議題
6. Applications 相關應用
7. Challenge and future work 挑戰與未來

What is *Bioinformatics*?

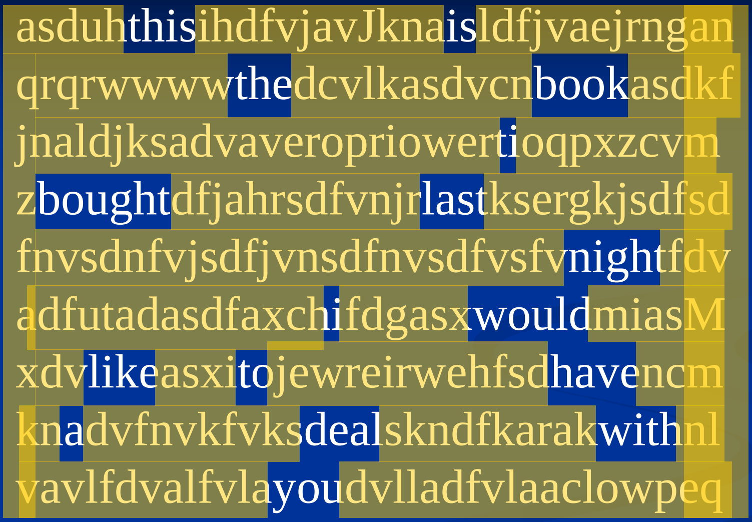
- 生物與電腦兩種科學之結合
- 利用電腦發展資料庫與分析工具，以協助分子生物學家之研究

…………………….The staggering volume of molecular data and its cryptic and subtle patterns have led to an absolute requirement for computerized databases and analysis tools…….. …………. new approaches to deal with the volume and complexity of data, and in providing researchers with better access to analysis and computing tools ………………… ………… advance understanding of our genetic legacy and its role in health and disease………………

(abstracted from 'general introduction of NCBI')

Definition
by National Human Genome Research Institute

- Bioinformatics, as related to ***genetics*** and ***genomics***, is a scientific subdiscipline that involves using ***computer technology*** to ***collect, store, analyze and disseminate*** biological data and information, such as DNA and amino acid sequences or annotations about those sequences.
- Scientists and clinicians use databases that organize and index such biological information to increase our understanding of health and disease and, in certain cases, as part of medical care.



asduhthisihdfvjavJknaisldfjvaejrngan

qrqrwwwwthedcvlkasdvcnbookasdkf

jnaldjksadvaveropriowertioqpxzcvm

zboughtdfjahrsdfvnjrlastksergkjsdfsd

fnvsdnfvjsdfjvnsdfnvsdfvsfvnightfdv

adfutadasdfaxchifdgasxwouldmiasM

xdvlikeasxitojewreirwehfsdhavencm

knadvfnvkfvksdealskndfkarakwithnl

vavlfdvalfvlayoudvlladfvlaaclowpeq

deoxyribonucleic acid (DNA) in chromosome can be represented as 4 characters: A, T, C, G

TTAGCTAGCCGCGC TGCGACGTTAACGT

GATCTTAGCTAGCG AGCAAGATCCTGAT

GGCATTATCTTCCG TCGGATTCTAGAGG

GAGGCGCGATACGG CGGAGGCGACGATG

AGGCGCGATATCTC CGCGTATATGCGAC

TCTTATCTTCTTTC GTTAACGTAGCAAG

GGAGGCGCGATATC ATCCTGAGATTCTA

TCTCTTATCTTCCG GAGGCGGAGGCGCT

GAGGCGCGATATCT ATCTTCCGGAGGCG

CTCTTATCTTCTGA CGATACGGAGGCGC

CGTTAACGTACGGA GATATCTCTCTTAG

GGCGCGATATCTCT CTAGCCGCGCGATC

CTTATCTTCTTAGC TTAGCTAGCGGGCA

TAGCCGCGCGATCT TCATGCTATTCGAG

TAGCTAGCGGGCAT GCGCTTAGCTAGCG

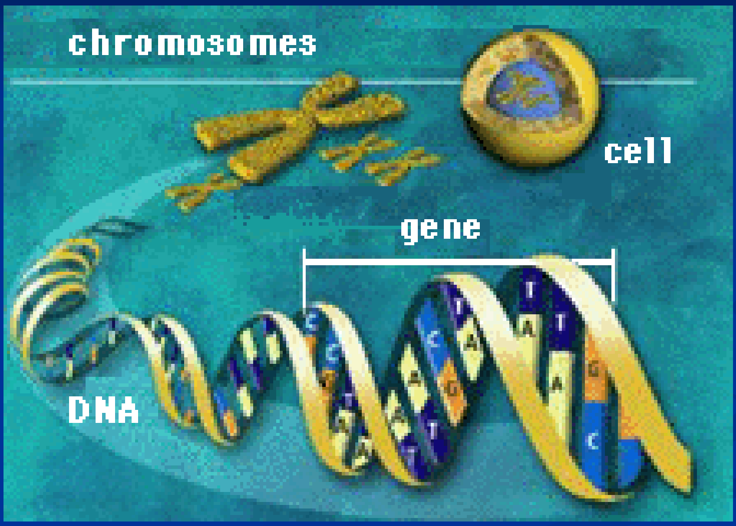
CATGCTCGGAGGCG GGCATDOGCTATTC

CGATATCTCTCTTA GGATTCTAGAGGCG

Multi-disciplinary Research Group
跨多重領域研究

- Computer scientists
- Molecular biologists
- Mathematicians
- Biochemists
- Research physicians

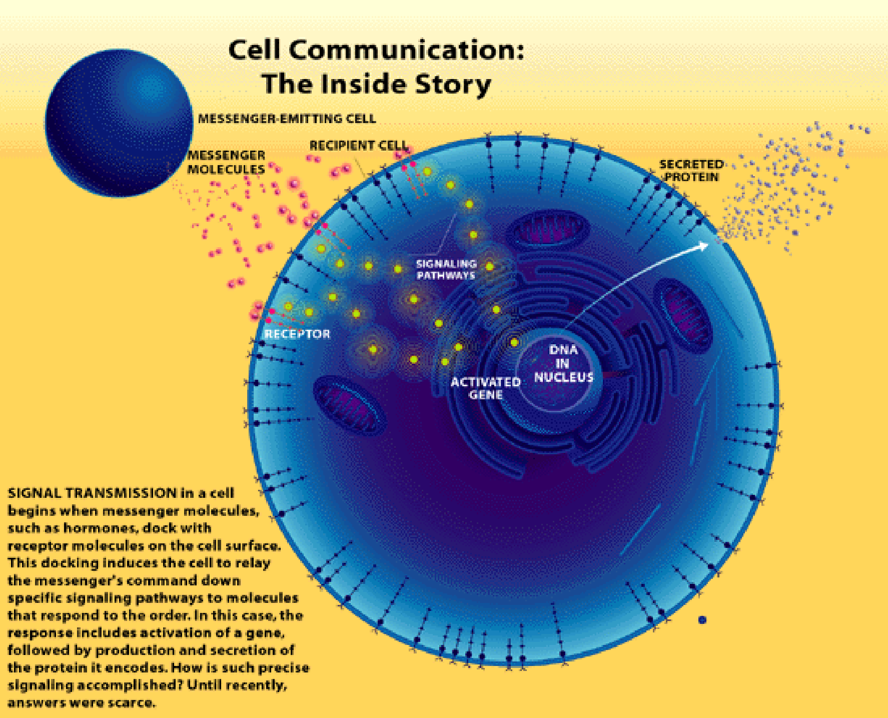
Molecular Genetics



Genes, DNA and Chromosomes explained

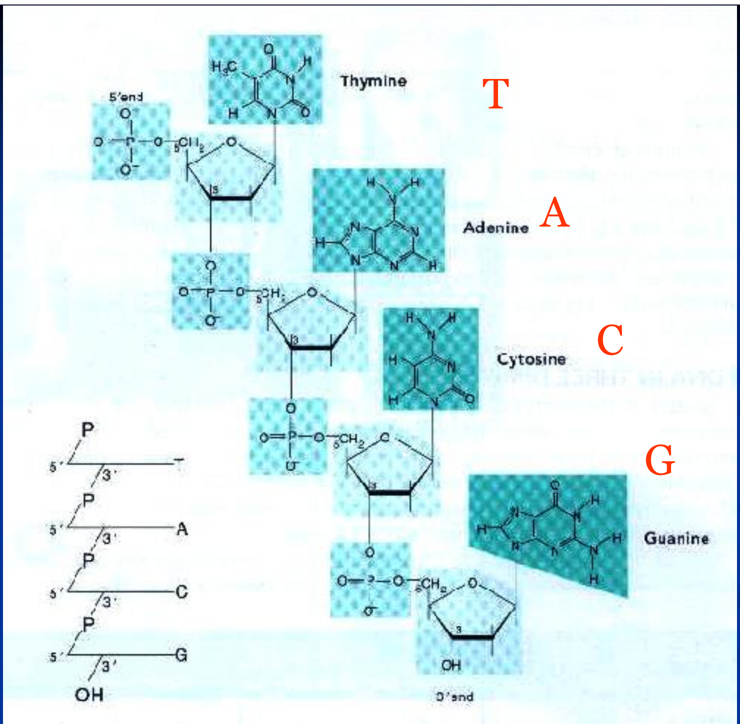
- https://www.youtube.com/watch?v=hywRdDVR76A&ab_channel=ScienceExplained

---



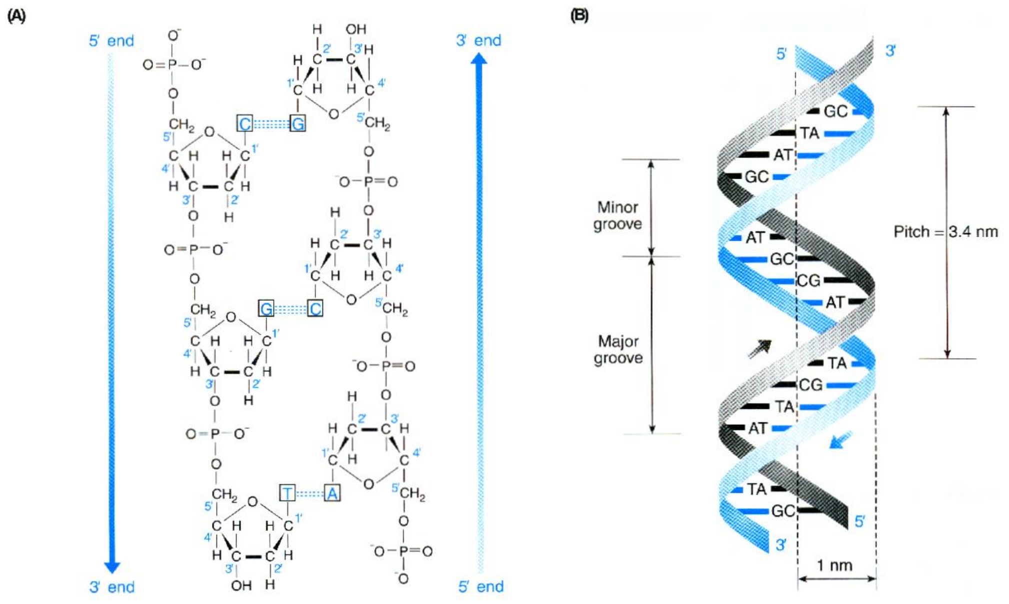
Cell Communication: The Inside Story

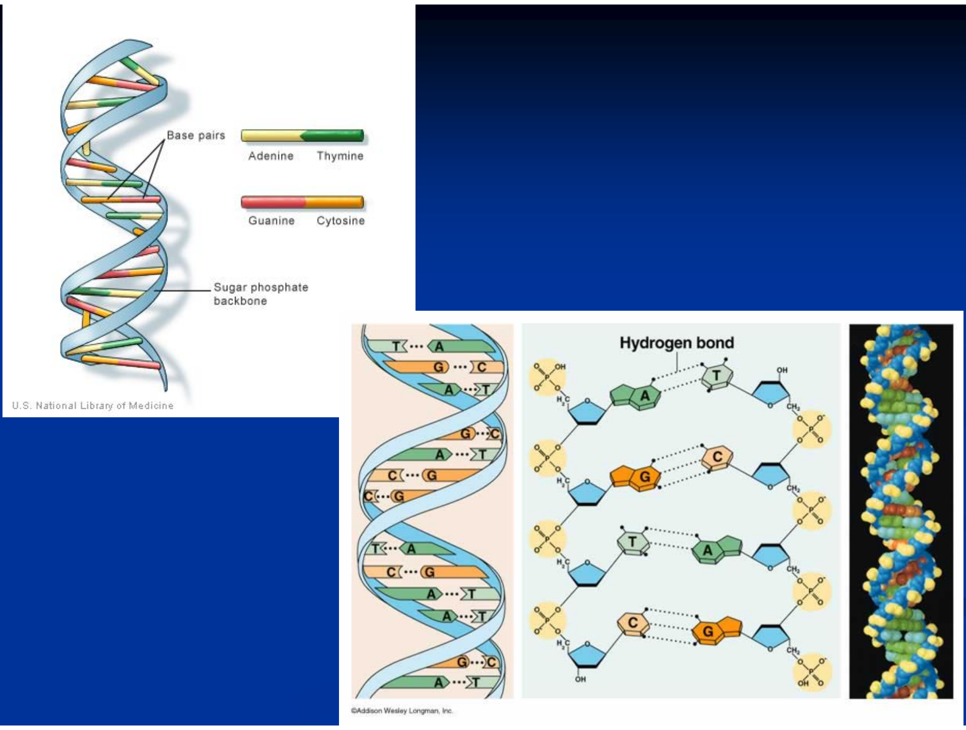
SIGNAL TRANSMISSION in a cell begins when messenger molecules, such as hormones, dock with receptor molecules on the cell surface. This docking induces the cell to relay the messenger's command down specific signaling pathways to molecules that respond to the order. In this case, the response includes activation of a gene, followed by production and secretion of the protein it encodes. How is such precise signaling accomplished? Until recently, answers were scarce.



4 types of
DNA molecules

Double Strand





U.S. National Library of Medicine

©Addison Wesley Longman, Inc.

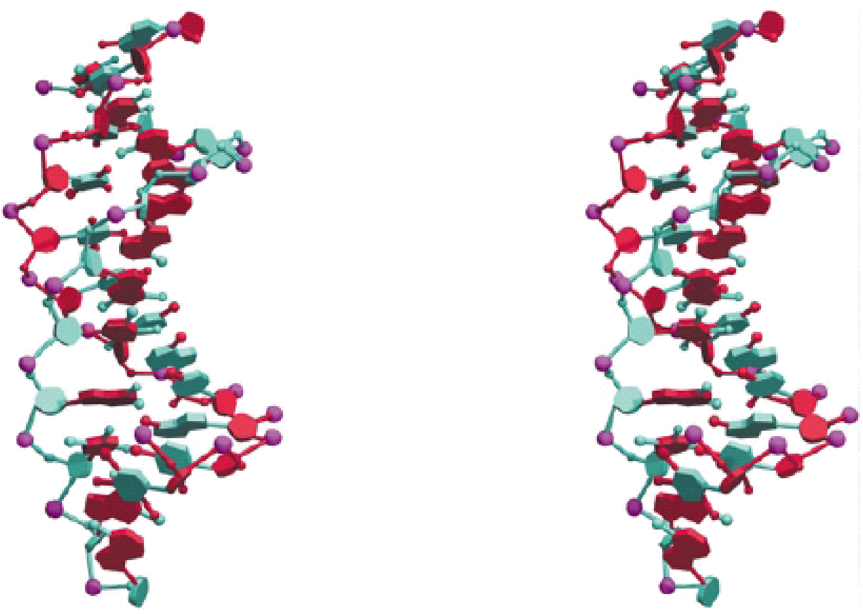
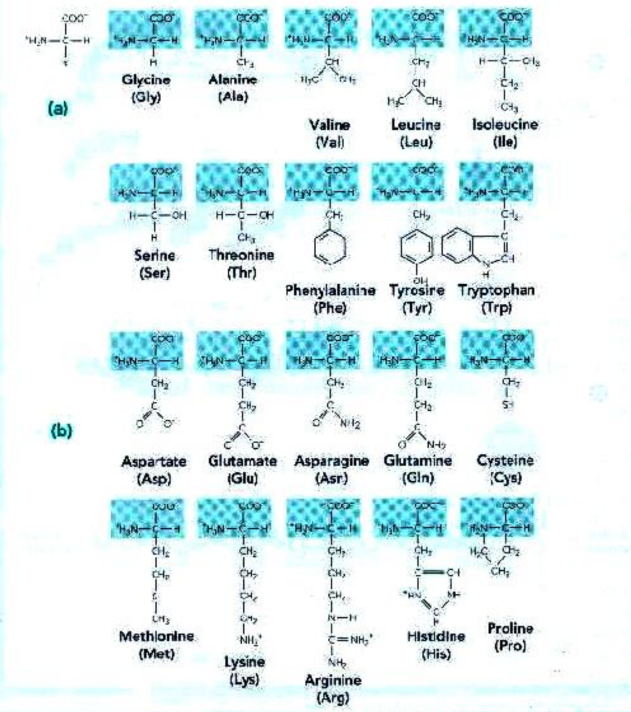
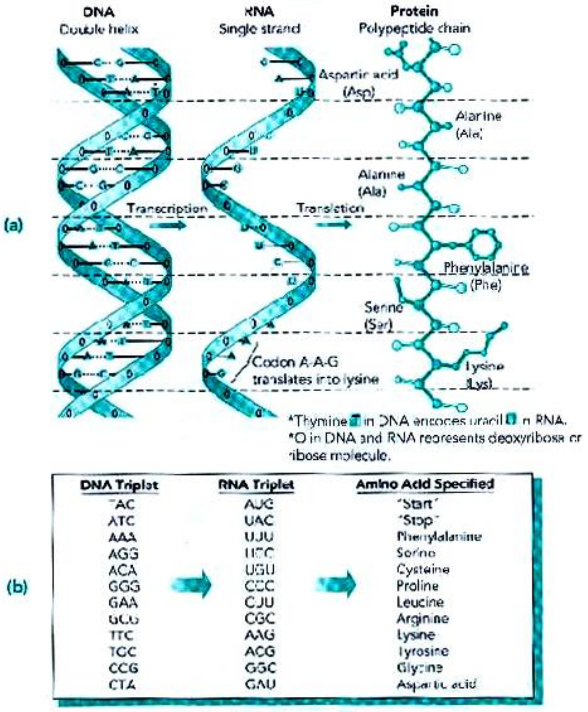


Plate I Double-helix of DNA. (See page 5.)



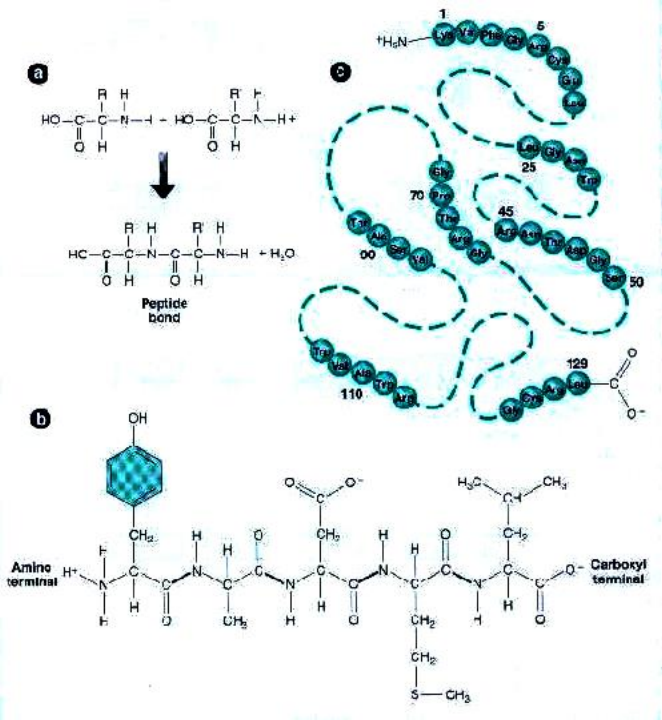
20 kinds of amino acids (胺基酸)

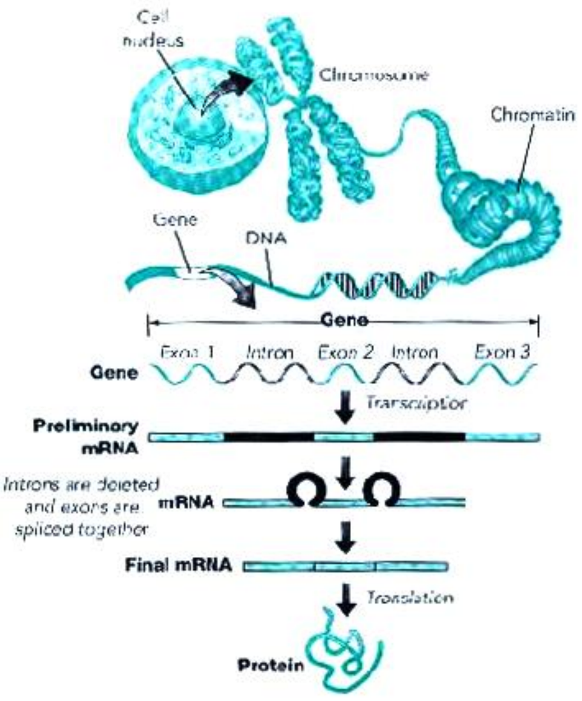


*Thymine in DNA encodes uracil in RNA.
*O in DNA and RNA represents deoxyribose or ribose molecule.

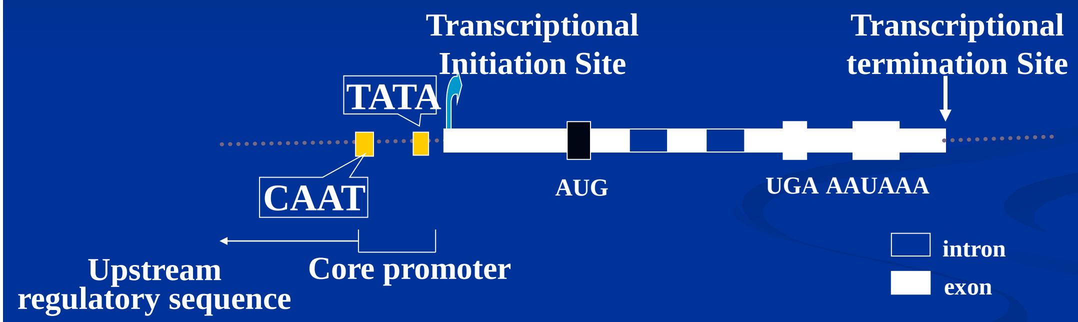
DNA Triplet	RNA Triplet	Amino Acid Specified
TAC	AUG	"Start"
ATC	UAC	"Stop"
AAA	UUU	Phenylalanine
AGG	UCC	Serine
ACA	UGU	Cysteine
GGG	CCC	Proline
GAA	CUU	Leucine
GCG	CGC	Arginine
TTC	AAG	Lysine
TGC	ACG	Tyrosine
CCG	GGC	Glycine
CTA	GAU	Aspartic acid

From DNA to Protein: Transcription & Translation


a
Peptide bond
b
Amino terminal
Carboxyl terminal
c
1
5
25
45
50
70
90
110
129


Cell nucleus
Chromosome
Chromatin
Gene
DNA
Gene
Exon 1
Intron
Exon 2
Intron
Exon 3
Gene
Transcription
Preliminary mRNA
Introns are deleted and exons are spliced together
mRNA
Final mRNA
Translation
Protein

Gene Structure



Transcriptional Initiation Site

Transcriptional termination Site

TATA

CAAT

AUG

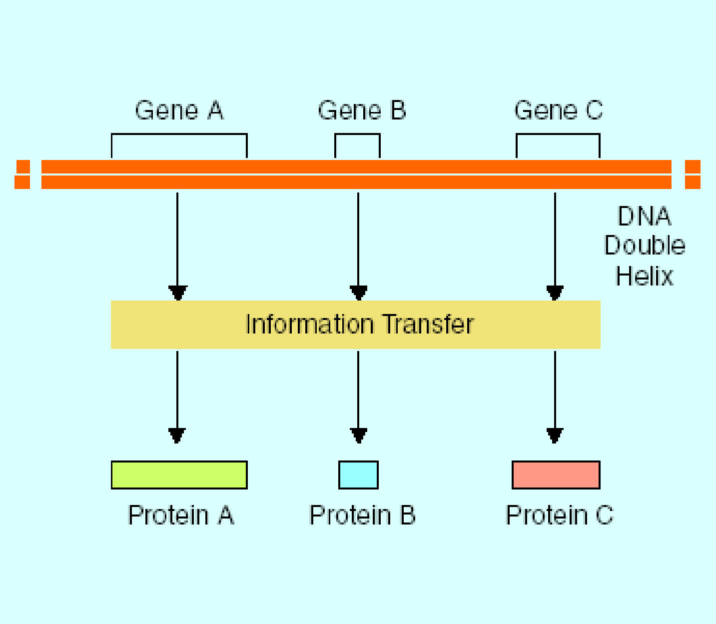
UGA AAUAAA

Upstream regulatory sequence

Core promoter

☐ intron

■ exon



Gene A

Gene B

Gene C

DNA Double Helix

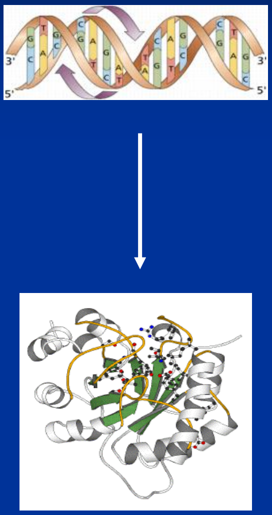
Information Transfer

Protein A

Protein B

Protein C

Gene: protein-coding DNA



DNA

transcription

mRNA

translation

Protein

CCTGAGCCAACTATTGATGAA

CCUGAGCCAACUAUUGAUGAA

PEPTIDE

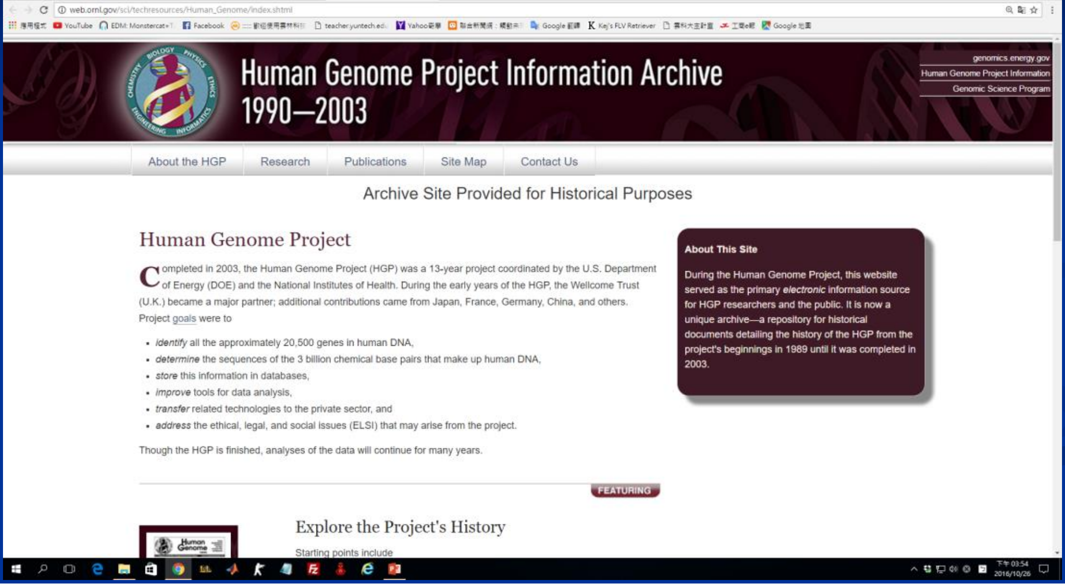
From DNA to protein - 3D film

https://www.youtube.com/watch?v=gG7uCskUOrA

人類基因體計畫
Human Genome Project

- 15-year-project; began in 1990, end in 2005
- Sponsored by the US Dept. of Energy and National Institute of Health; $3 billion budget
- 18 countries worldwide participated, with significant contributions from the Sanger Center in the United Kingdom and research centers in Germany, France, and Japan. China recently joined the effort.

Human Genome Project



Archive Site Provided for Historical Purposes

Human Genome Project

Completed in 2003, the Human Genome Project (HGP) was a 13-year project coordinated by the U.S. Department of Energy (DOE) and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others. Project goals were to

- *identify* all the approximately 20,500 genes in human DNA,
- *determine* the sequences of the 3 billion chemical base pairs that make up human DNA,
- *store* this information in databases,
- *improve* tools for data analysis,
- *transfer* related technologies to the private sector, and
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

Though the HGP is finished, analyses of the data will continue for many years.

About This Site

During the Human Genome Project, this website served as the primary *electronic* information source for HGP researchers and the public. It is now a unique archive—a repository for historical documents detailing the history of the HGP from the project's beginnings in 1989 until it was completed in 2003.

Explore the Project's History

Starting points include

Project Goals

- *Identify* all the approximate 20,500 genes in human DNA
- *Determine* the sequences of the 3 billion chemical base pairs that make up human DNA
- *Store* this information in databases
- *Develop* tools for data analysis
- *Transfer* related technologies to the private sector
- *Address* the ethical, legal, and social issues

Human Genome Project

Human Genome Project Information Archive
1990—2003

HGP RESEARCH

- HGP Research by Topic
  - Sequencing
  - Sequencing Technologies
  - Mapping
  - Bioinformatics
  - Functional and Comparative Genomics
  - Ethical, Legal, and Social Issues (ELSI)
- Data-Release Policies
- Landmark HGP Papers

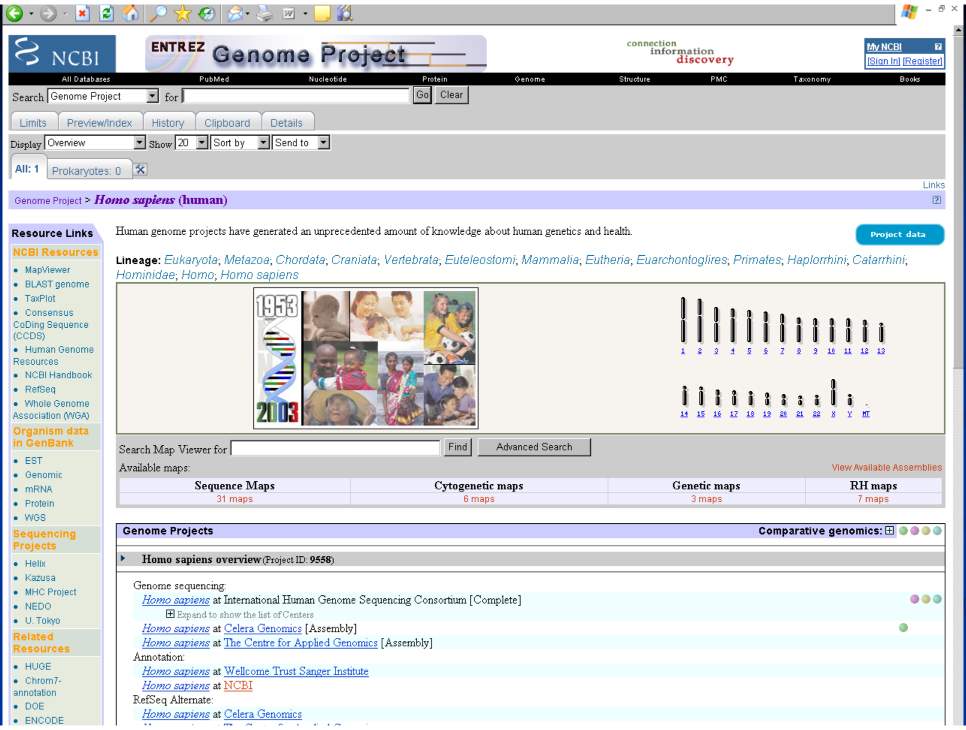
CURRENT DOE GENOMICS RESEARCH

The DOE Genomic Science Program uses microbial and plant genomic data, high-throughput analytical technologies, and modeling and simulation to develop a predictive understanding of biological systems behavior relevant to solving energy and environmental challenges including bioenergy production, environmental remediation, and climate stabilization.

DOE JOINT GENOME INSTITUTE

DOE JGI in Walnut Creek, California, provides state-of-the-science capabilities for genome sequencing and analysis. With more than 1100 worldwide collaborators on active projects, JGI is the preeminent facility for sequencing plants, microbes, and microbial communities that are foundational to energy and environmental research.

Explore the Project's History



Genome Project > ***Homo sapiens* (human)**

Human genome projects have generated an unprecedented amount of knowledge about human genetics and health.

Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens*

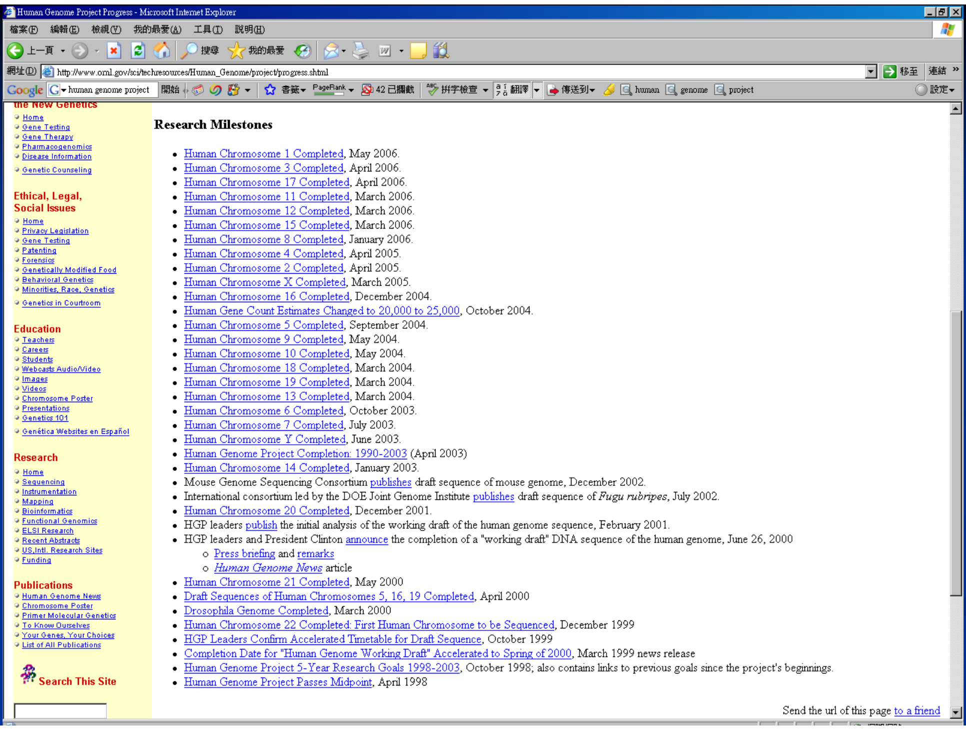
Available maps:

Sequence Maps	Cytogenetic maps	Genetic maps	RH maps
31 maps	6 maps	3 maps	7 maps

Genome Projects

Homo sapiens overview (Project ID: **9558**)

Genome sequencing:
Homo sapiens at International Human Genome Sequencing Consortium [Complete]
⊞ Expand to show the list of Centers
Homo sapiens at Celera Genomics [Assembly]
Homo sapiens at The Centre for Applied Genomics [Assembly]
Annotation:
Homo sapiens at Wellcome Trust Sanger Institute
Homo sapiens at NCBI
RefSeq Alternate:
Homo sapiens at Celera Genomics



Research Milestones

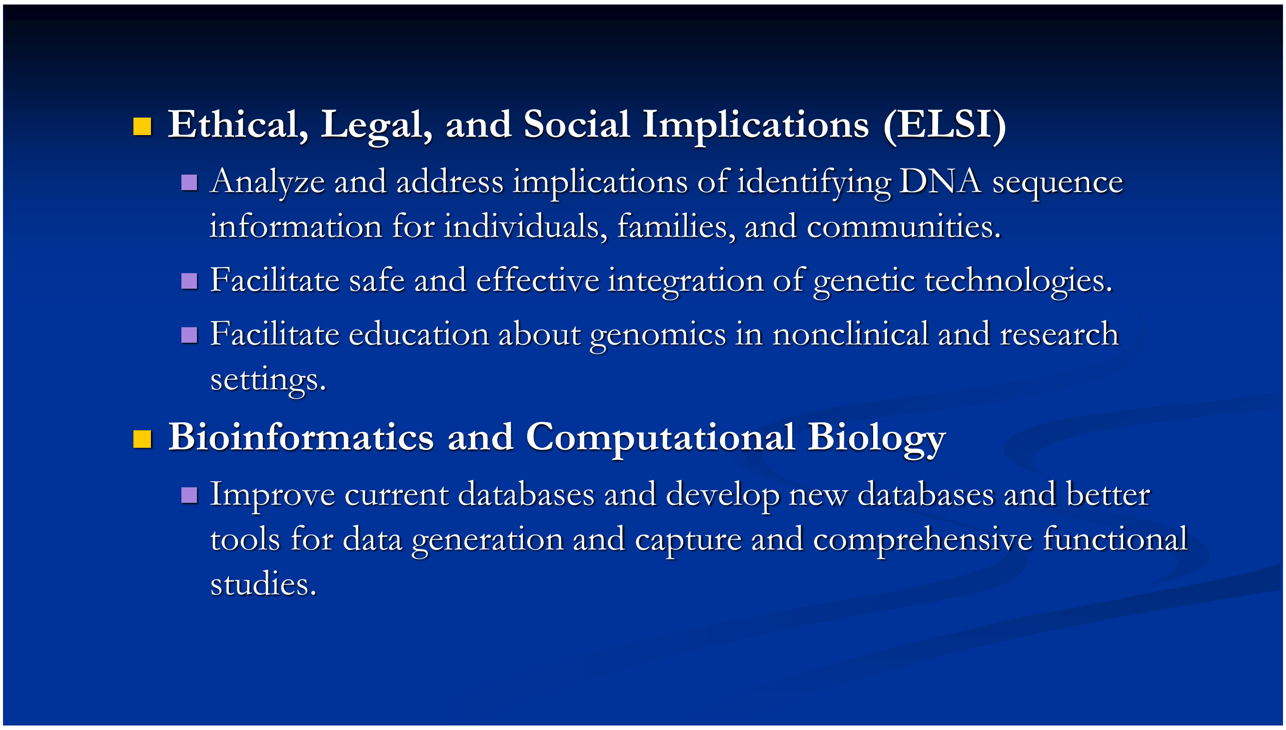
- Human Chromosome 1 Completed, May 2006.
- Human Chromosome 3 Completed, April 2006.
- Human Chromosome 17 Completed, April 2006.
- Human Chromosome 11 Completed, March 2006.
- Human Chromosome 12 Completed, March 2006.
- Human Chromosome 15 Completed, March 2006.
- Human Chromosome 8 Completed, January 2006.
- Human Chromosome 4 Completed, April 2005.
- Human Chromosome 2 Completed, April 2005.
- Human Chromosome X Completed, March 2005.
- Human Chromosome 16 Completed, December 2004.
- Human Gene Count Estimates Changed to 20,000 to 25,000, October 2004.
- Human Chromosome 5 Completed, September 2004.
- Human Chromosome 9 Completed, May 2004.
- Human Chromosome 10 Completed, May 2004.
- Human Chromosome 18 Completed, March 2004.
- Human Chromosome 19 Completed, March 2004.
- Human Chromosome 13 Completed, March 2004.
- Human Chromosome 6 Completed, October 2003.
- Human Chromosome 7 Completed, July 2003.
- Human Chromosome Y Completed, June 2003.
- Human Genome Project Completion: 1990-2003 (April 2003)
- Human Chromosome 14 Completed, January 2003.
- Mouse Genome Sequencing Consortium publishes draft sequence of mouse genome, December 2002.
- International consortium led by the DOE Joint Genome Institute publishes draft sequence of *Fugu rubripes*, July 2002.
- Human Chromosome 20 Completed, December 2001.
- HGP leaders publish the initial analysis of the working draft of the human genome sequence, February 2001.
- HGP leaders and President Clinton announce the completion of a "working draft" DNA sequence of the human genome, June 26, 2000
  - Press briefing and remarks
  - *Human Genome News* article
- Human Chromosome 21 Completed, May 2000
- Draft Sequences of Human Chromosomes 5, 16, 19 Completed, April 2000
- Drosophila Genome Completed, March 2000
- Human Chromosome 22 Completed: First Human Chromosome to be Sequenced, December 1999
- HGP Leaders Confirm Accelerated Timetable for Draft Sequence, October 1999
- Completion Date for "Human Genome Working Draft" Accelerated to Spring of 2000, March 1999 news release
- Human Genome Project 5-Year Research Goals 1998-2003, October 1998; also contains links to previous goals since the project's beginnings.
- Human Genome Project Passes Midpoint, April 1998

Send the url of this page to a friend

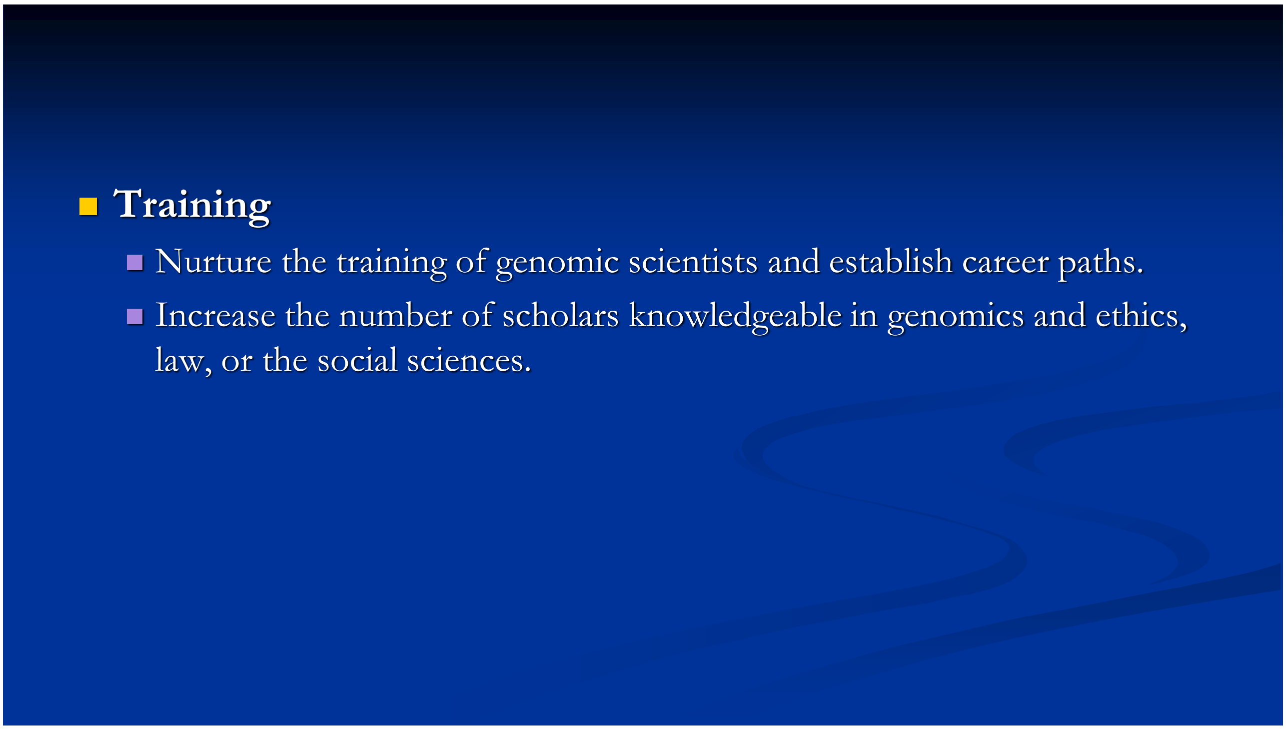
Highlights of the New 5-Year Plan for the U.S. Human Genome Project

- **Human DNA Sequencing**
  - **Generate a working draft of 90% of the genome (2001).**
  - **Obtain a complete, high-quality genomic sequence (2003).**
  - **Make all data publicly available.**

- **Sequencing Technology**
  - Create a long-term, sustainable sequencing capacity by improving current technology and developing highly efficient novel technologies.
- **Sequence Variation**
  - Develop technologies for rapid identification of DNA sequence variants.
- **Functional Genomics**
  - Expand support for current approaches and innovative technologies.
- **Comparative Genomics**
  - Obtain complete genomic sequences for *C. elegans* (1998), *Drosophila* (2002), and mouse (2008).



- **Ethical, Legal, and Social Implications (ELSI)**
  - Analyze and address implications of identifying DNA sequence information for individuals, families, and communities.
  - Facilitate safe and effective integration of genetic technologies.
  - Facilitate education about genomics in nonclinical and research settings.
- **Bioinformatics and Computational Biology**
  - Improve current databases and develop new databases and better tools for data generation and capture and comprehensive functional studies.



- **Training**
  - Nurture the training of genomic scientists and establish career paths.
  - Increase the number of scholars knowledgeable in genomics and ethics, law, or the social sciences.

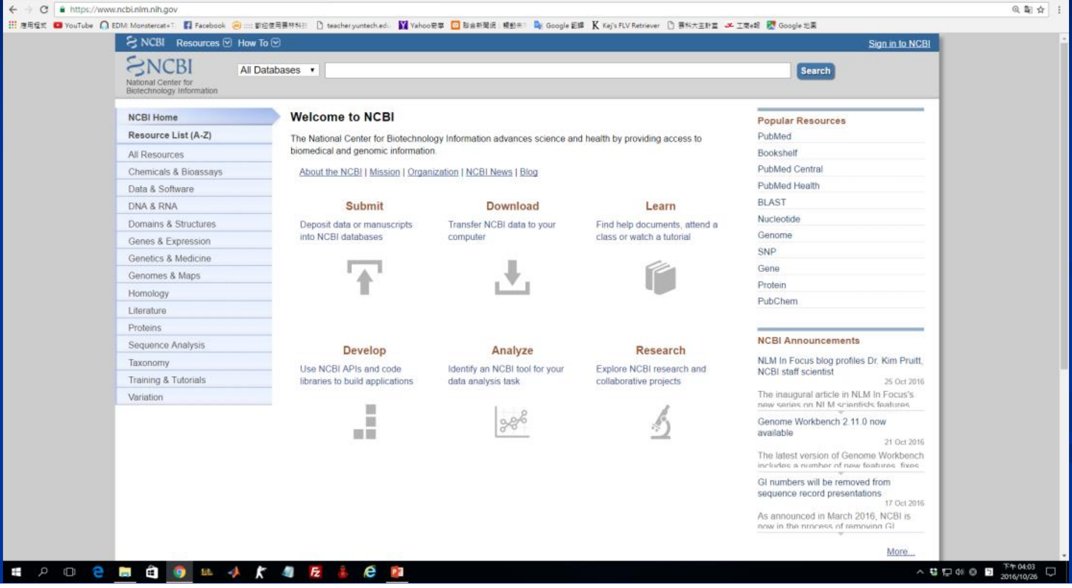
Genome Scientists' To-Do List

- Correct errors and proofread
- Fill tens of thousands of gaps
- Sequence the 7 percent
- Finish finding all the genes that make proteins
- Find the non-protein-making genes
- Discover the regulatory sequences
- Untangle the genes' intricate interactions with other molecules
- Identify gene functions

Databases & information retrieval

- GeneBank – Maintained by National Center for Biotechnology Information (NCBI), USA.
  - It is divided into several sections with sequences grouped according to species.
  - Searches can be made by keywords or by sequence.
  - **http://www.ncbi.nlm.nih.gov/**

NCBI website



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 | NCBI News | Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

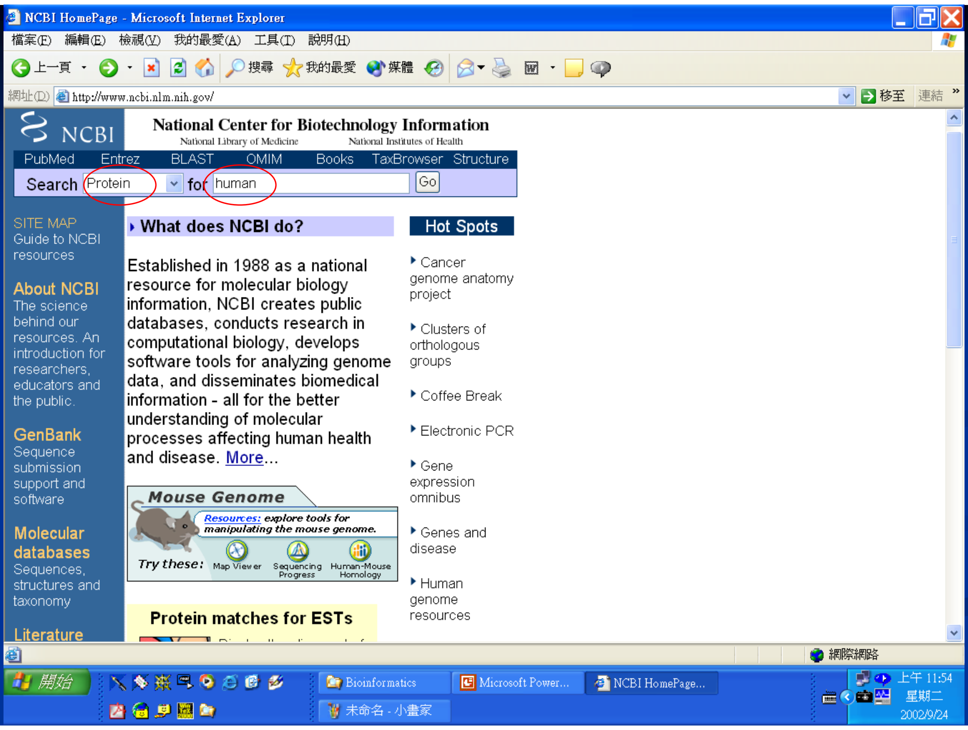
Research
Explore NCBI research and collaborative projects

NCBI Announcements

NLM In Focus blog profiles Dr. Kim Pruitt, NCBI staff scientist
25 Oct 2016
The inaugural article in NLM In Focus's new series on NLM scientists features

Genome Workbench 2.11.0 now available
21 Oct 2016
The latest version of Genome Workbench includes a number of new features, fixes

GI numbers will be removed from sequence record presentations
17 Oct 2016
As announced in March 2016, NCBI is now in the process of removing GI



National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

Search Protein for human

SITE MAP
Guide to NCBI resources

About NCBI
The science behind our resources. An introduction for researchers, educators and the public.

GenBank
Sequence submission support and software

Molecular databases
Sequences, structures and taxonomy

Literature

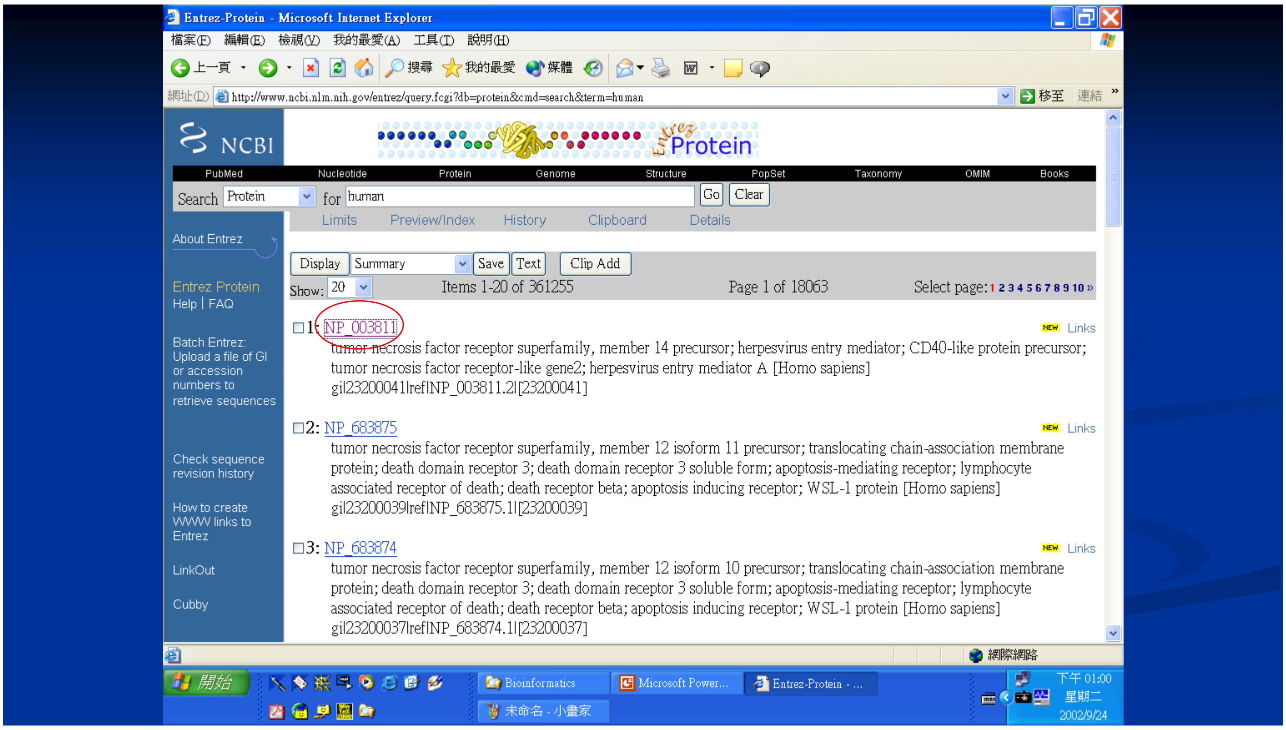
What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

Resources: explore tools for manipulating the mouse genome.

Protein matches for ESTs

- Cancer genome anatomy project
- Clusters of orthologous groups
- Coffee Break
- Electronic PCR
- Gene expression omnibus
- Genes and disease
- Human genome resources



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein&cmd=search&term=human

Search Protein for human

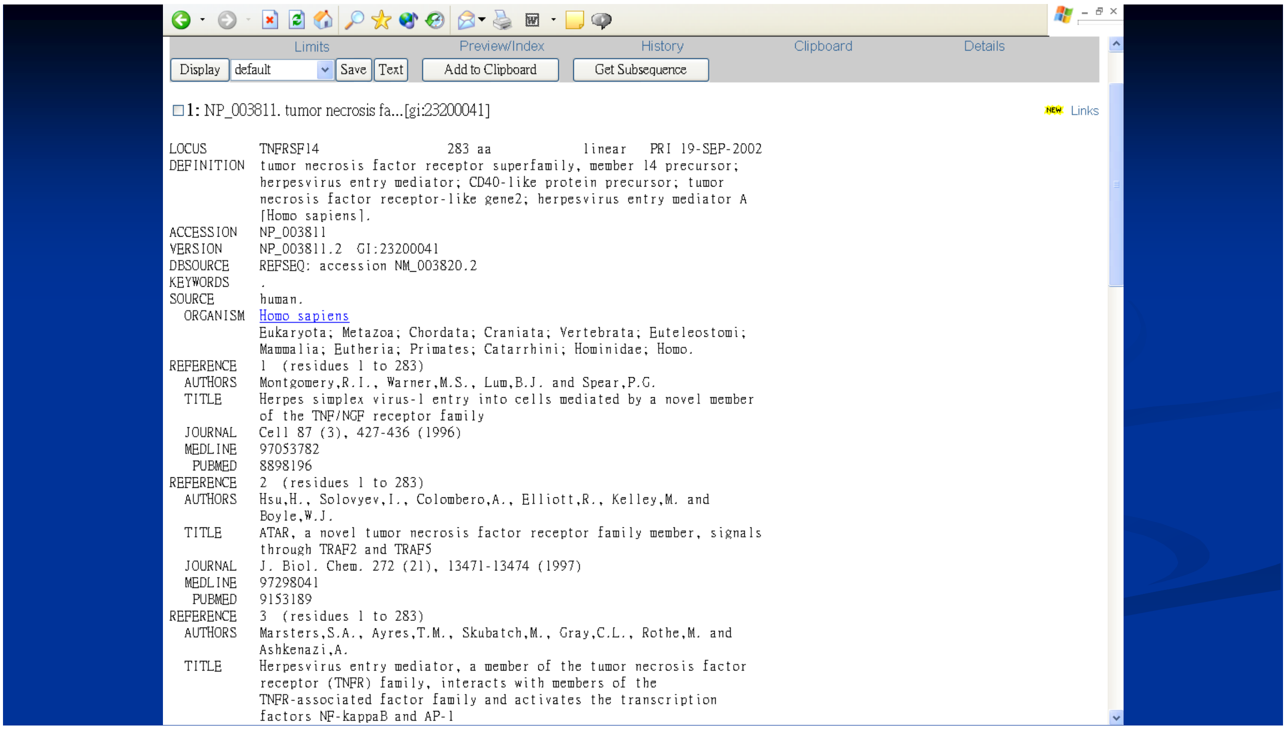
Display Summary

Show: 20 Items 1-20 of 361255 Page 1 of 18063

☐1: NP_003811
tumor necrosis factor receptor superfamily, member 14 precursor; herpesvirus entry mediator; CD40-like protein precursor; tumor necrosis factor receptor-like gene2; herpesvirus entry mediator A [Homo sapiens]
gi|23200041|ref|NP_003811.2|[23200041]

☐2: NP_683875
tumor necrosis factor receptor superfamily, member 12 isoform 11 precursor; translocating chain-association membrane protein; death domain receptor 3; death domain receptor 3 soluble form; apoptosis-mediating receptor; lymphocyte associated receptor of death; death receptor beta; apoptosis inducing receptor; WSL-1 protein [Homo sapiens]
gi|23200039|ref|NP_683875.1|[23200039]

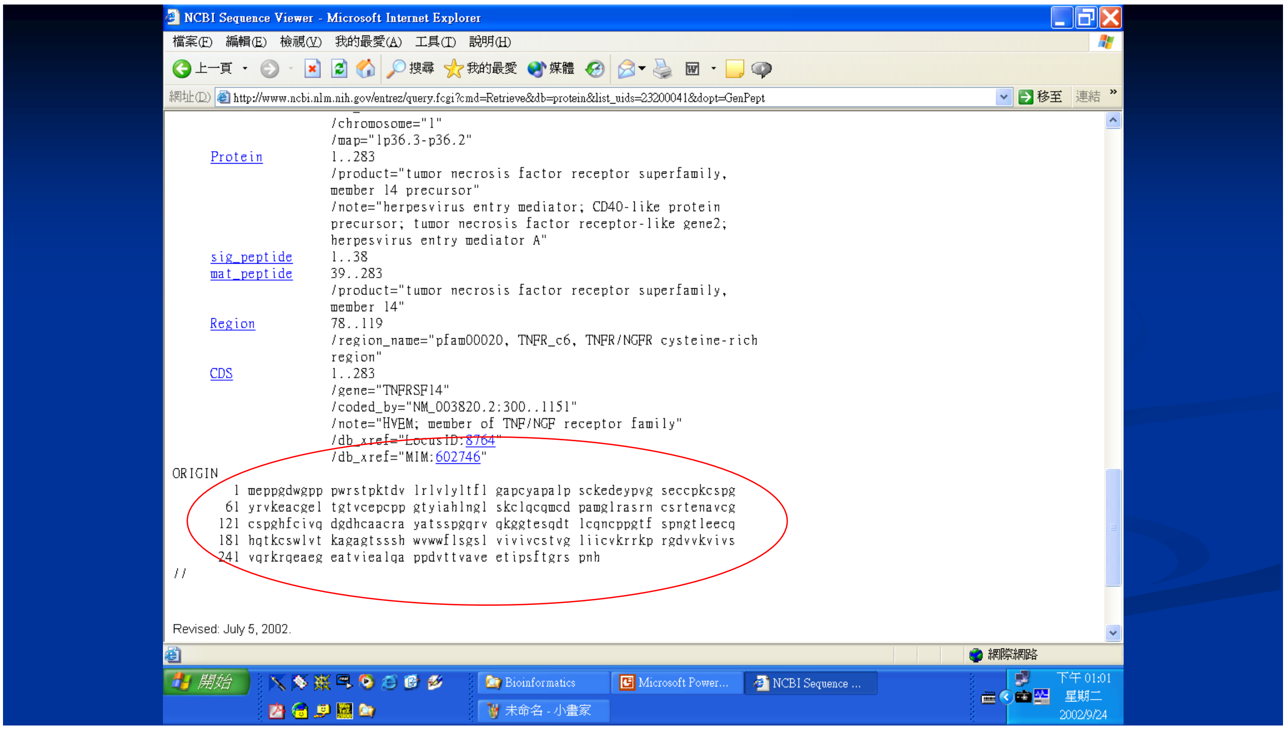
☐3: NP_683874
tumor necrosis factor receptor superfamily, member 12 isoform 10 precursor; translocating chain-association membrane protein; death domain receptor 3; death domain receptor 3 soluble form; apoptosis-mediating receptor; lymphocyte associated receptor of death; death receptor beta; apoptosis inducing receptor; WSL-1 protein [Homo sapiens]
gi|23200037|ref|NP_683874.1|[23200037]



Display default

☐1: NP_003811. tumor necrosis fa...[gi:23200041]

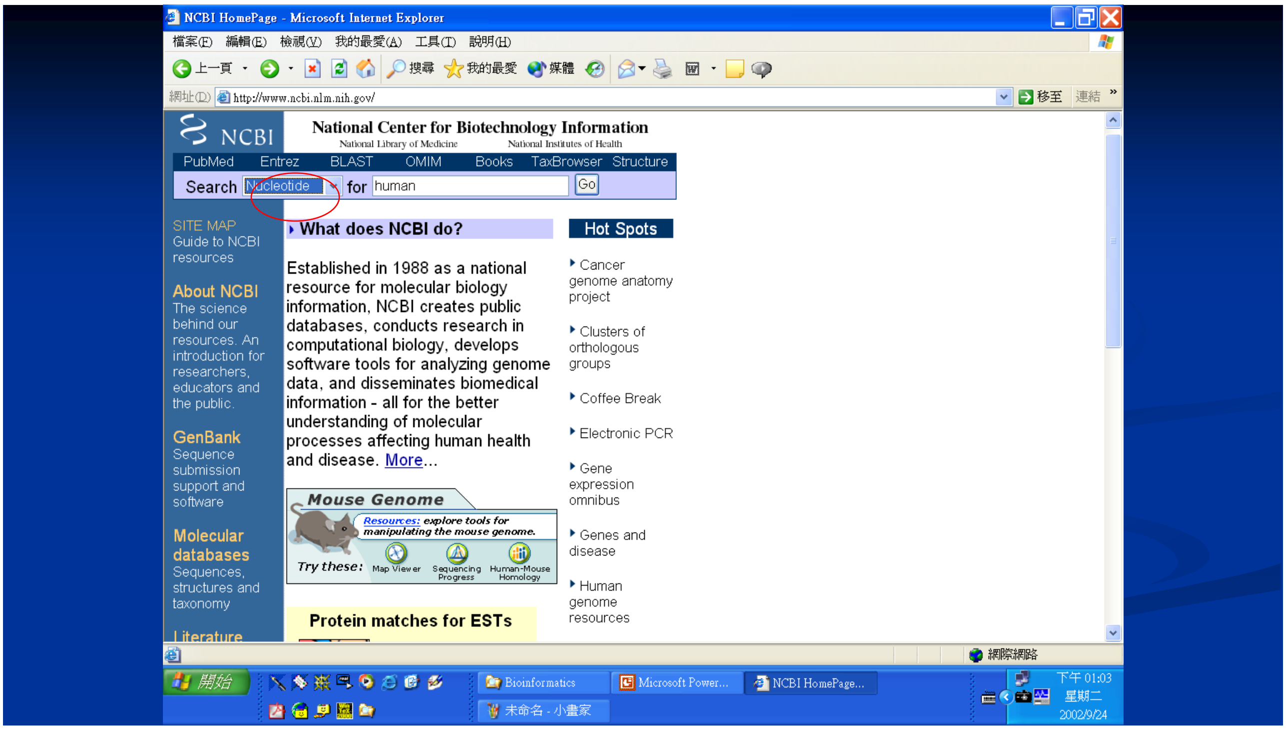
```
LOCUS       TNFRSF14                 283 aa            linear   PRI 19-SEP-2002
DEFINITION  tumor necrosis factor receptor superfamily, member 14 precursor;
            herpesvirus entry mediator; CD40-like protein precursor; tumor
            necrosis factor receptor-like gene2; herpesvirus entry mediator A
            [Homo sapiens].
ACCESSION   NP_003811
VERSION     NP_003811.2  GI:23200041
DBSOURCE    REFSEQ: accession NM_003820.2
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 283)
  AUTHORS   Montgomery,R.I., Warner,M.S., Lum,B.J. and Spear,P.G.
  TITLE     Herpes simplex virus-1 entry into cells mediated by a novel member
            of the TNF/NGF receptor family
  JOURNAL   Cell 87 (3), 427-436 (1996)
  MEDLINE   97053782
   PUBMED   8898196
REFERENCE   2  (residues 1 to 283)
  AUTHORS   Hsu,H., Solovyev,I., Colombero,A., Elliott,R., Kelley,M. and
            Boyle,W.J.
  TITLE     ATAR, a novel tumor necrosis factor receptor family member, signals
            through TRAF2 and TRAF5
  JOURNAL   J. Biol. Chem. 272 (21), 13471-13474 (1997)
  MEDLINE   97298041
   PUBMED   9153189
REFERENCE   3  (residues 1 to 283)
  AUTHORS   Marsters,S.A., Ayres,T.M., Skubatch,M., Gray,C.L., Rothe,M. and
            Ashkenazi,A.
  TITLE     Herpesvirus entry mediator, a member of the tumor necrosis factor
            receptor (TNFR) family, interacts with members of the
            TNFR-associated factor family and activates the transcription
            factors NF-kappaB and AP-1
```



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=23200041&dopt=GenPept

```
                     /chromosome="1"
                     /map="1p36.3-p36.2"
     Protein         1..283
                     /product="tumor necrosis factor receptor superfamily,
                     member 14 precursor"
                     /note="herpesvirus entry mediator; CD40-like protein
                     precursor; tumor necrosis factor receptor-like gene2;
                     herpesvirus entry mediator A"
     sig_peptide     1..38
     mat_peptide     39..283
                     /product="tumor necrosis factor receptor superfamily,
                     member 14"
     Region          78..119
                     /region_name="pfam00020, TNFR_c6, TNFR/NGFR cysteine-rich
                     region"
     CDS             1..283
                     /gene="TNFRSF14"
                     /coded_by="NM_003820.2:300..1151"
                     /note="HVEM; member of TNF/NGF receptor family"
                     /db_xref="LocusID:8764"
                     /db_xref="MIM:602746"
ORIGIN
        1 meppgdwgpp pwrstpktdv lrlvlyltfl gapcyapalp sckedeypvg seccpkcspg
       61 yrvkeacgel tgtvcepcpp gtyiahlngl skclqcqmcd pamglrasrn csrtenavcg
      121 cspghfcivq dgdhcaacra yatsspgqrv qkggtesqdt lcqncppgtf spngtleecq
      181 hqtkcswlvt kagagtsssh wvwwflsgsl vivivcstvg liicvkrrkp rgdvvkvivs
      241 vqrkrqeaeg eatviealqa ppdvttvave etipsftgrs pnh
//
```

Revised: July 5, 2002.



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

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

Search Nucleotide for human

SITE MAP
Guide to NCBI resources

About NCBI
The science behind our resources. An introduction for researchers, educators and the public.

GenBank
Sequence submission support and software

Molecular databases
Sequences, structures and taxonomy

Literature

What does NCBI do?

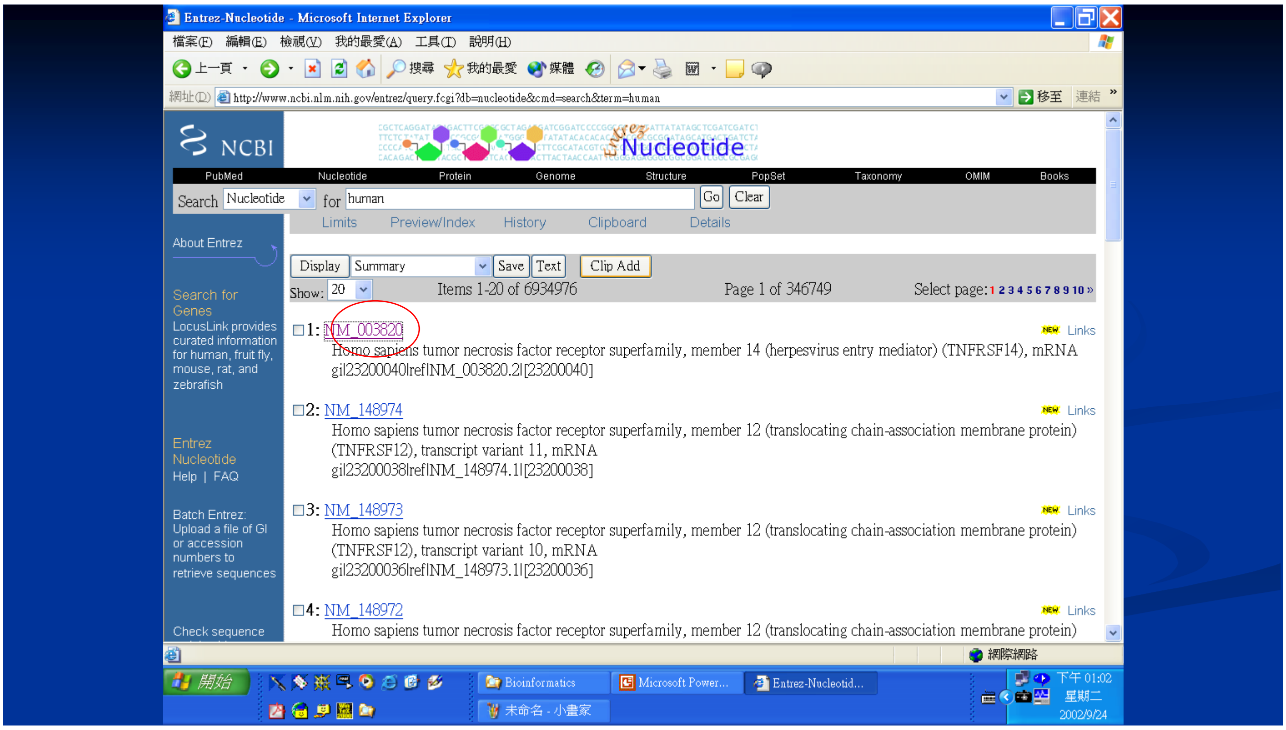
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

Resources: explore tools for manipulating the mouse genome.

Protein matches for ESTs

Hot Spots

- Cancer genome anatomy project
- Clusters of orthologous groups
- Coffee Break
- Electronic PCR
- Gene expression omnibus
- Genes and disease
- Human genome resources



Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Entrez Nucleotide
Help | FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve sequences

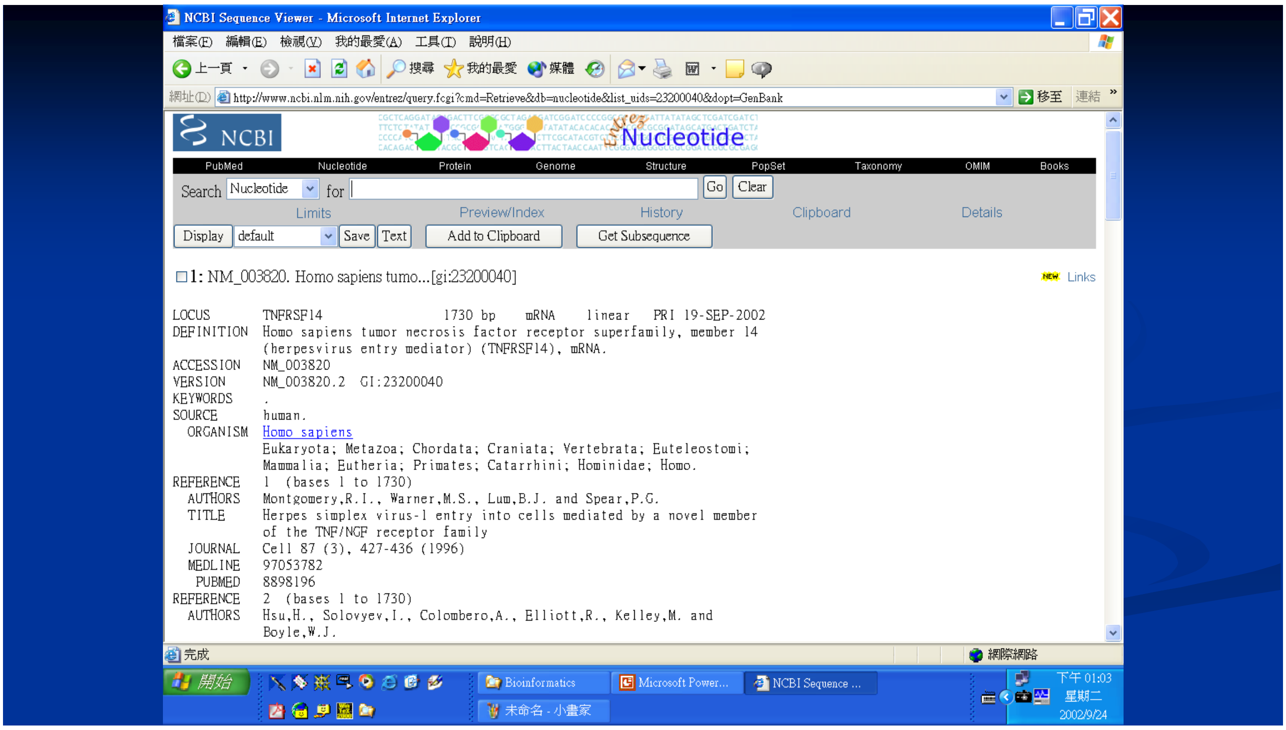
Check sequence

☐1: NM_003820
Homo sapiens tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator) (TNFRSF14), mRNA
gi|23200040|ref|NM_003820.2|[23200040]

☐2: NM_148974
Homo sapiens tumor necrosis factor receptor superfamily, member 12 (translocating chain-association membrane protein) (TNFRSF12), transcript variant 11, mRNA
gi|23200038|ref|NM_148974.1|[23200038]

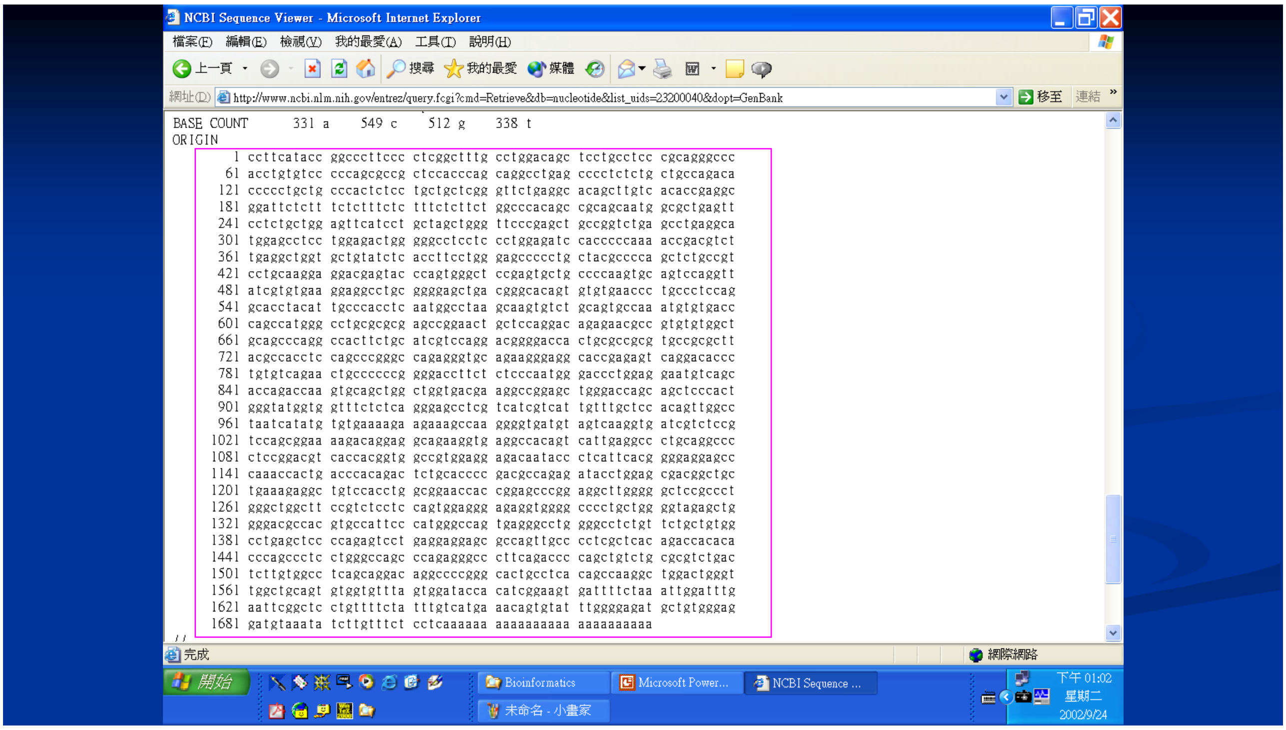
☐3: NM_148973
Homo sapiens tumor necrosis factor receptor superfamily, member 12 (translocating chain-association membrane protein) (TNFRSF12), transcript variant 10, mRNA
gi|23200036|ref|NM_148973.1|[23200036]

☐4: NM_148972
Homo sapiens tumor necrosis factor receptor superfamily, member 12 (translocating chain-association membrane protein)



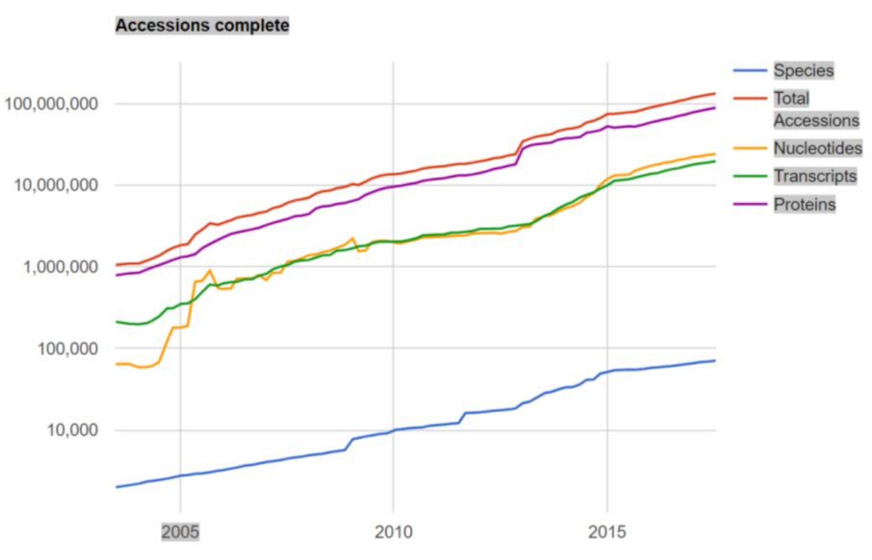
☐1: NM_003820. Homo sapiens tumo...[gi:23200040]

```
LOCUS       TNFRSF14                1730 bp    mRNA    linear   PRI 19-SEP-2002
DEFINITION  Homo sapiens tumor necrosis factor receptor superfamily, member 14
            (herpesvirus entry mediator) (TNFRSF14), mRNA.
ACCESSION   NM_003820
VERSION     NM_003820.2  GI:23200040
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1730)
  AUTHORS   Montgomery,R.I., Warner,M.S., Lum,B.J. and Spear,P.G.
  TITLE     Herpes simplex virus-1 entry into cells mediated by a novel member
            of the TNF/NGF receptor family
  JOURNAL   Cell 87 (3), 427-436 (1996)
  MEDLINE   97053782
   PUBMED   8898196
REFERENCE   2  (bases 1 to 1730)
  AUTHORS   Hsu,H., Solovyev,I., Colombero,A., Elliott,R., Kelley,M. and
            Boyle,W.J.
```

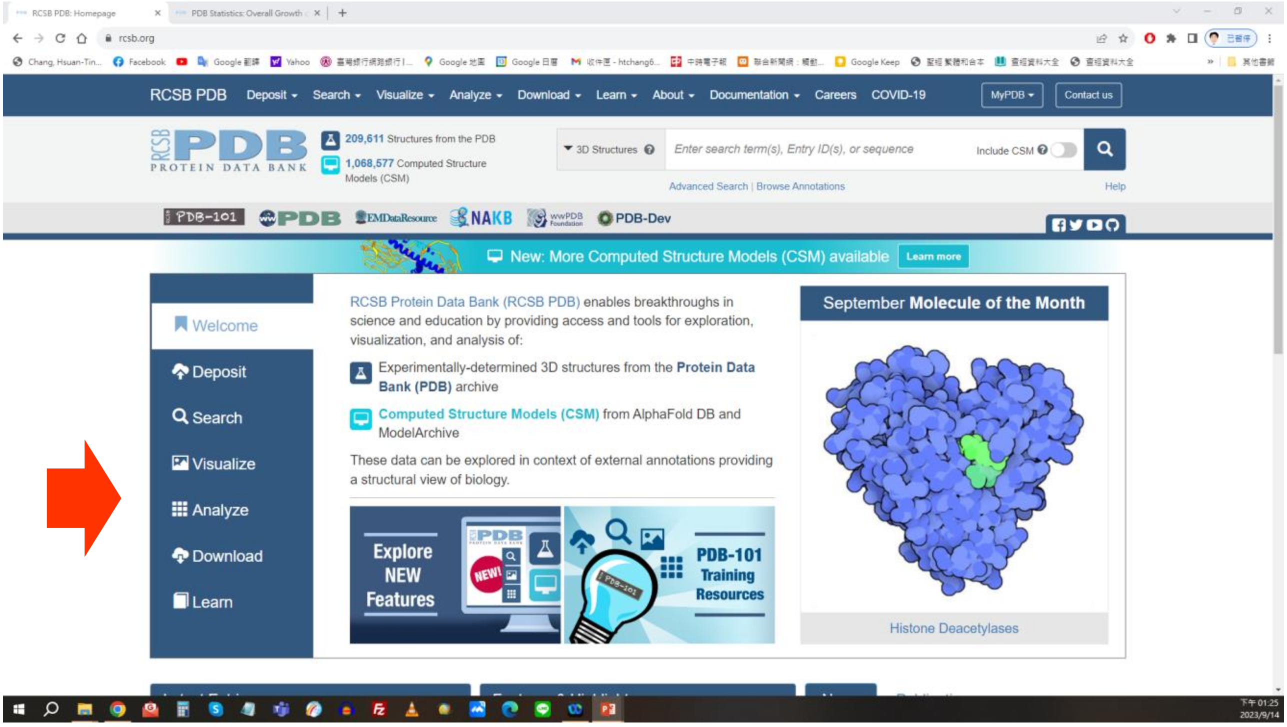


```
BASE COUNT      331 a     549 c     512 g     338 t
ORIGIN
        1 ccttcatacc ggcccttccc ctcggctttg cctggacagc tcctgcctcc cgcagggccc
       61 acctgtgtcc cccagcgccg ctccacccag caggcctgag cccctctctg ctgccagaca
      121 ccccctgctg cccactctcc tgctgctcgg gttctgaggc acagcttgtc acaccgaggc
      181 ggattctctt tctctttctc tttctcttct ggcccacagc cgcagcaatg gcgctgagtt
      241 cctctgctgg agttcatcct gctagctggg ttcccgagct gccggtctga gcctgaggca
      301 tggagcctcc tggagactgg gggcctcctc cctggagatc cacccccaaa accgacgtct
      361 tgaggctggt gctgtatctc accttcctgg gagccccctg ctacgcccca gctctgccgt
      421 cctgcaagga ggacgagtac ccagtgggct ccgagtgctg ccccaagtgc agtccaggtt
      481 atcgtgtgaa ggaggcctgc ggggagctga cgggcacagt gtgtgaaccc tgccctccag
      541 gcacctacat tgcccacctc aatggcctaa gcaagtgtct gcagtgccaa atgtgtgacc
      601 cagccatggg cctgcgcgcg agccggaact gctccaggac agagaacgcc gtgtgtggct
      661 gcagcccagg ccacttctgc atcgtccagg acggggacca ctgcgccgcg tgccgcgctt
      721 acgccacctc cagcccgggc cagagggtgc agaagggagg caccgagagt caggacaccc
      781 tgtgtcagaa ctgccccccg gggaccttct ctcccaatgg gaccctggag gaatgtcagc
      841 accagaccaa gtgcagctgg ctggtgacga aggccggagc tgggaccagc agctcccact
      901 gggtatggtg gtttctctca gggagcctcg tcatcgtcat tgtttgctcc acagttggcc
      961 taatcatatg tgtgaaaaga agaaagccaa ggggtgatgt agtcaaggtg atcgtctccg
     1021 tccagcggaa aagacaggag gcagaaggtg aggccacagt cattgaggcc ctgcaggccc
     1081 ctccggacgt caccacggtg gccgtggagg agacaatacc ctcattcacg gggaggagcc
     1141 caaaccactg acccacagac tctgcacccc gacgccagag atacctggag cgacggctgc
     1201 tgaaagaggc tgtccacctg gcggaaccac cggagcccgg aggcttgggg gctccgccct
     1261 gggctggctt ccgtctcctc cagtggaggg agaggtgggg cccctgctgg ggtagagctg
     1321 gggacgccac gtgccattcc catgggccag tgagggcctg gggcctctgt tctgctgtgg
     1381 cctgagctcc ccagagtcct gaggaggagc gccagttgcc cctcgctcac agaccacaca
     1441 cccagccctc ctgggccagc ccagagggcc cttcagaccc cagctgtctg cgcgtctgac
     1501 tcttgtggcc tcagcaggac aggccccggg cactgcctca cagccaaggc tggactgggt
     1561 tggctgcagt gtggtgttta gtggatacca catcggaagt gattttctaa attggatttg
     1621 aattcggctc ctgttttcta tttgtcatga aacagtgtat ttggggagat gctgtgggag
     1681 gatgtaaata tcttgtttct cctcaaaaaa aaaaaaaaaa aaaaaaaaaa
//
```

RefSeq Growth Statistics



- EMBL: The European Molecular Biology Laboratory.
  - https://www.embl.org/
- GDB: Genome Database
  - https://www.ncbi.nlm.nih.gov/genome/
- PDB: The Protein Data Bank, a repository of 3-D structure of proteins.
  - https://www.wwpdb.org/



209,611 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

New: More Computed Structure Models (CSM) available

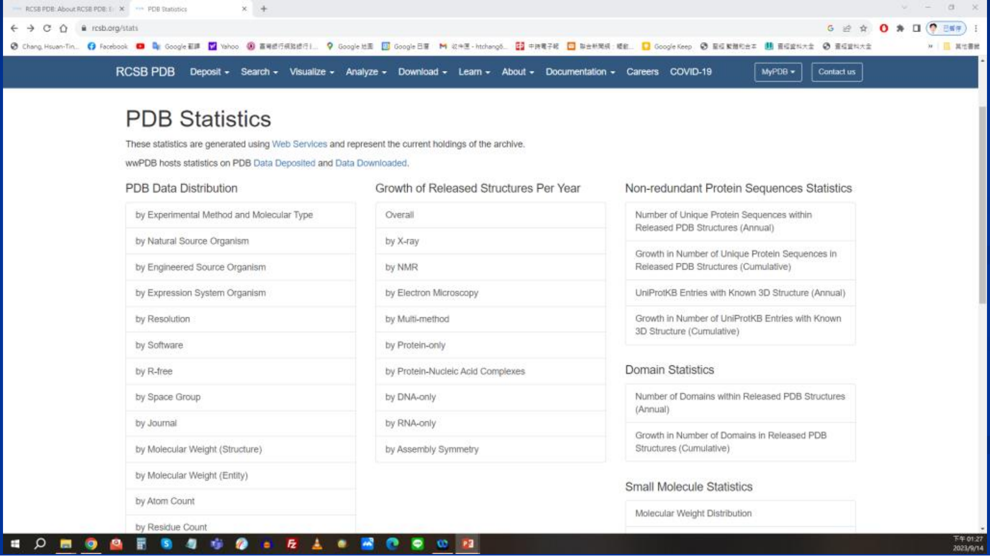
RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive
- **Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

RCSB PDB Statistics

- https://www.rcsb.org/stats



RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

PDB Statistics

These statistics are generated using Web Services and represent the current holdings of the archive.

wwPDB hosts statistics on PDB Data Deposited and Data Downloaded.

PDB Data Distribution

- by Experimental Method and Molecular Type
- by Natural Source Organism
- by Engineered Source Organism
- by Expression System Organism
- by Resolution
- by Software
- by R-free
- by Space Group
- by Journal
- by Molecular Weight (Structure)
- by Molecular Weight (Entity)
- by Atom Count
- by Residue Count

Growth of Released Structures Per Year

- Overall
- by X-ray
- by NMR
- by Electron Microscopy
- by Multi-method
- by Protein-only
- by Protein-Nucleic Acid Complexes
- by DNA-only
- by RNA-only
- by Assembly Symmetry

Non-redundant Protein Sequences Statistics

- Number of Unique Protein Sequences within Released PDB Structures (Annual)
- Growth in Number of Unique Protein Sequences in Released PDB Structures (Cumulative)
- UniProtKB Entries with Known 3D Structure (Annual)
- Growth in Number of UniProtKB Entries with Known 3D Structure (Cumulative)

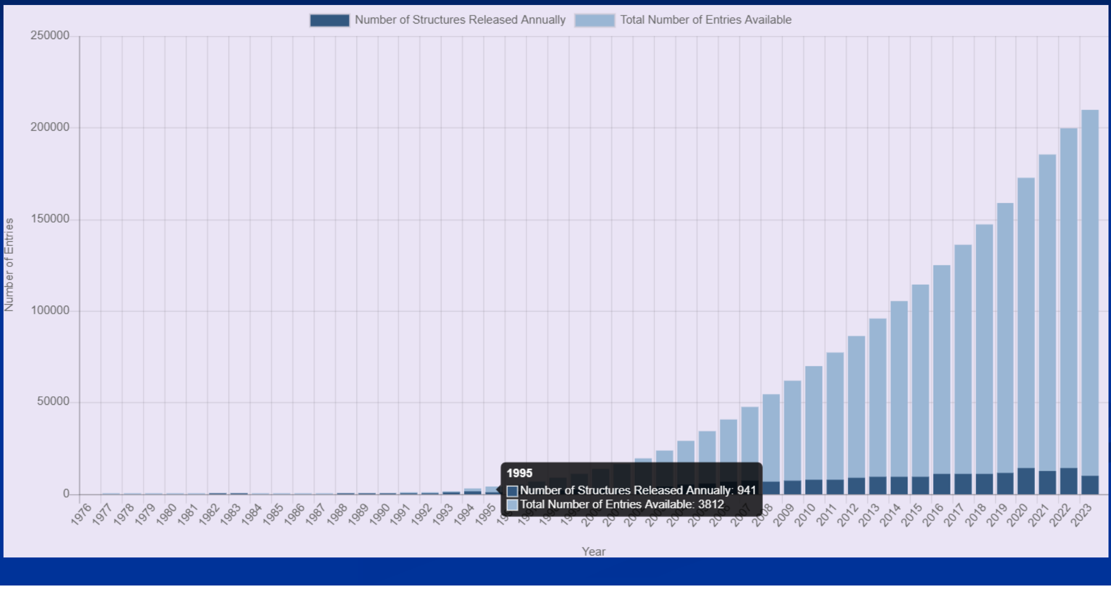
Domain Statistics

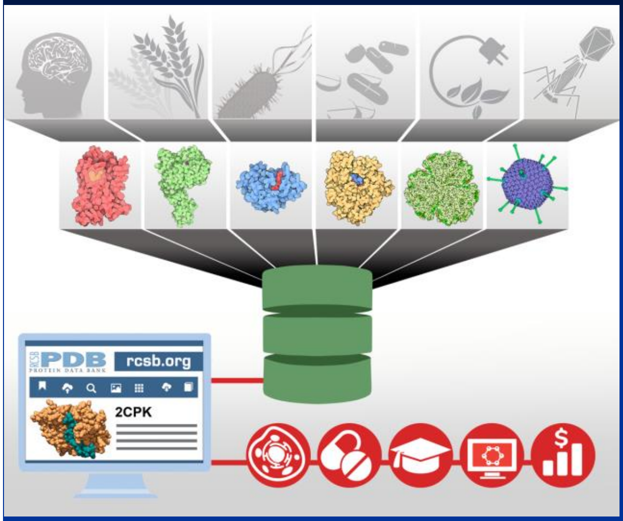
- Number of Domains within Released PDB Structures (Annual)
- Growth in Number of Domains in Released PDB Structures (Cumulative)

Small Molecule Statistics

- Molecular Weight Distribution

Protein Data Bank Overall Growth of Released Structures Per Year



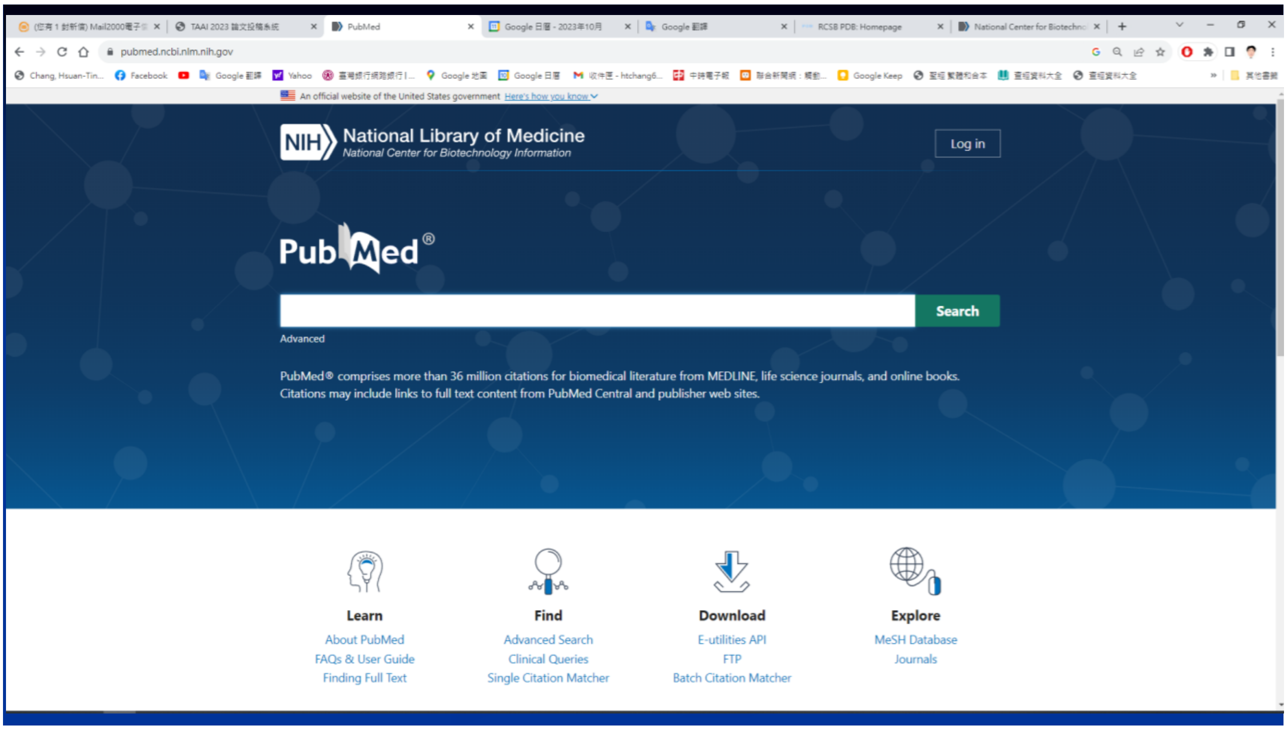


RCSB PDB (RCSB.org) is the US data center for the global Protein Data Bank (PDB) archive of 3D structure data for large biological molecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology.

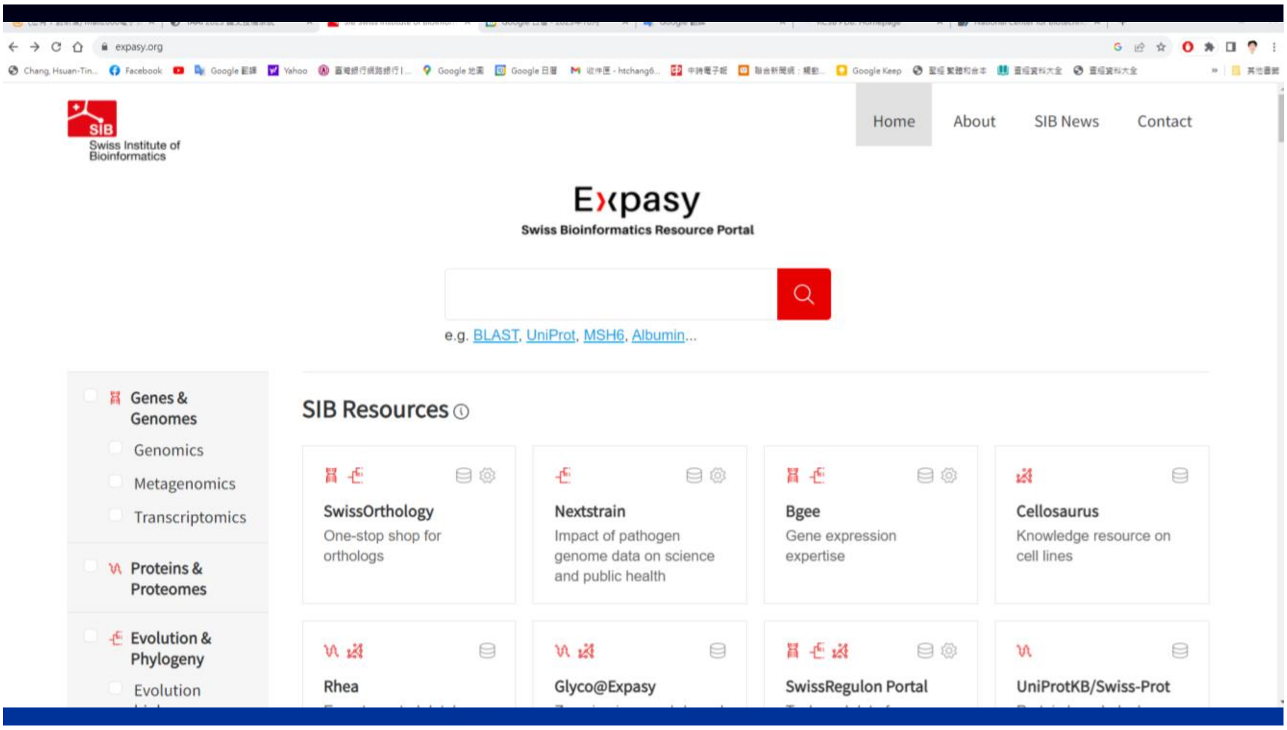
PubMed Central® (PMC)

PubMed Central

- A free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM).
- In keeping with NLM's legislative mandate to collect and preserve the biomedical literature, PMC is part of the NLM collection, which also includes NLM's extensive print and licensed electronic journal holdings and supports contemporary biomedical and health care research and practice as well as future scholarship.



PubMed® comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.



ExPASy

- The ExPASy (the ***Expert Protein Analysis System***) World Wide Web server (http://www.expasy.org), is provided as a service to the life science community by a multidisciplinary team at the Swiss Institute of Bioinformatics (SIB).
- It provides access to a variety of databases and analytical tools dedicated to proteins and proteomics.
- ExPASy databases include SWISS-PROT and TrEMBL, SWISS-2DPAGE, PROSITE, ENZYME and the SWISS-MODEL repository.

Research Issues

- Sequence comparison & database search
- Fragment assembly of DNA
- Physical mapping of DNA
- Phylogenetic trees
- Genome rearrangement
- Molecular structure prediction

Global Research on Bioinformatics

Major Cartels in the World

- In Europe
  - EMBL @ https://www.embl.org/
  - EBI @ https://www.ebi.ac.uk/
- In North America
  - NCBI @ https://www.ncbi.nlm.nih.gov/
- In Asia
  - GenomeNet @ https://www.genome.ad.jp/about_genomenet/about.html

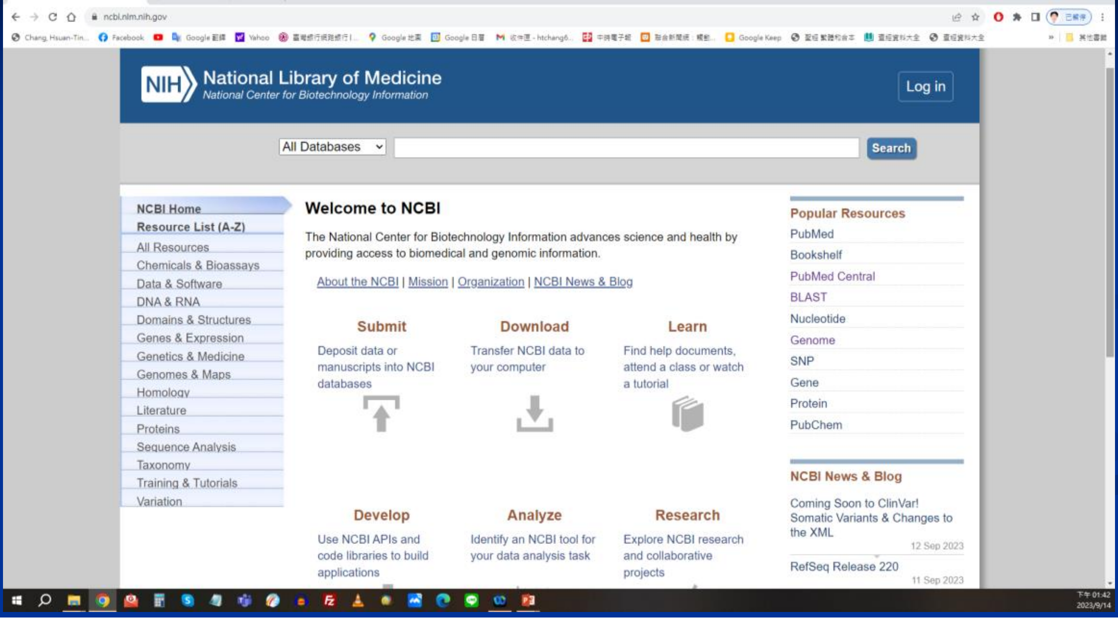
Global Research Projects on Bioinformatics

- Detection and analysis of gene organization
- Repeating sequence patterns
- Protein domains and structural elements
- Creation of a gene map of the human genome
- Analysis of effects of sequencing errors for database searching

Global Research Projects (cont.)

- Development of new algorithms for database searching and multiple sequence alignment
- Construction of non-redundant sequence databases
- Mathematical models for estimation of statistical significance of sequence similarity
- Establishment of microarray systems for screening disease genes

The National Center for Biotechnology Information



ncbi.nlm.nih.gov

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

All Databases | Search

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 | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

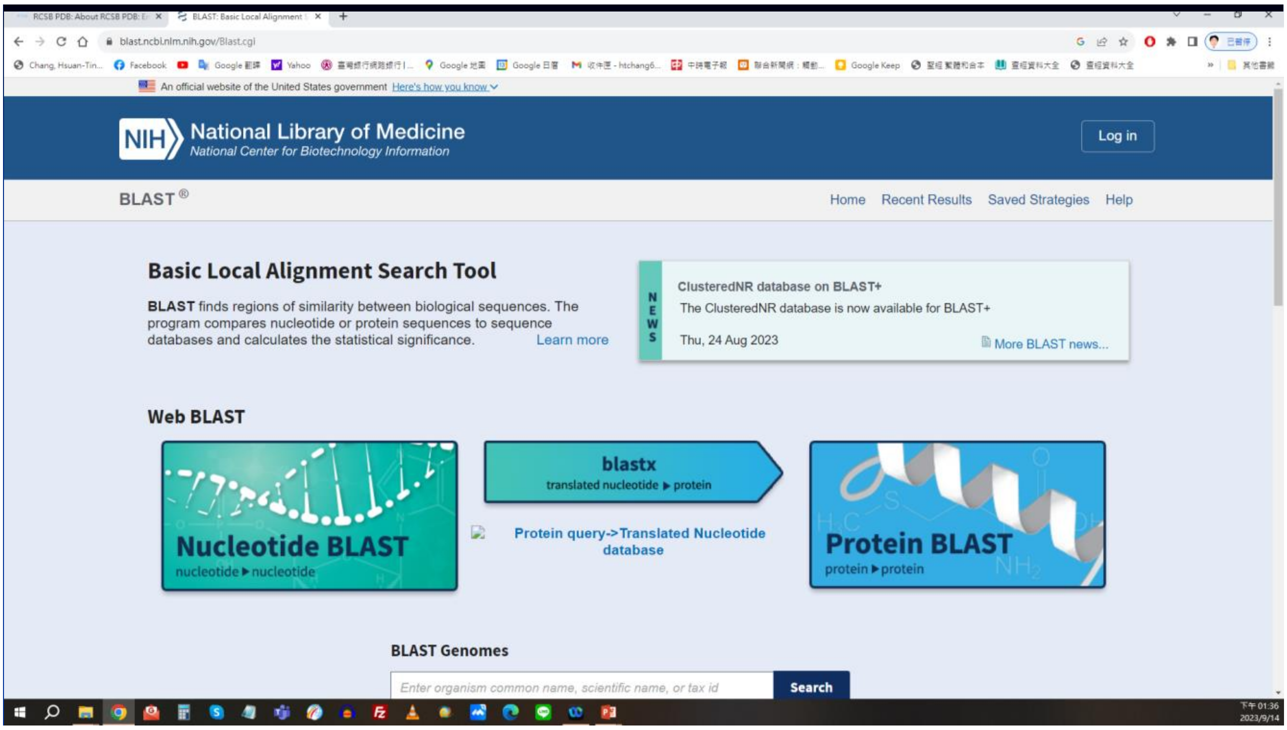
Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

NCBI News & Blog

Coming Soon to ClinVar! Somatic Variants & Changes to the XML
12 Sep 2023

RefSeq Release 220
11 Sep 2023



blast.ncbi.nlm.nih.gov/Blast.cgi

An official website of the United States government Here's how you know

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

Basic Local Alignment Search Tool

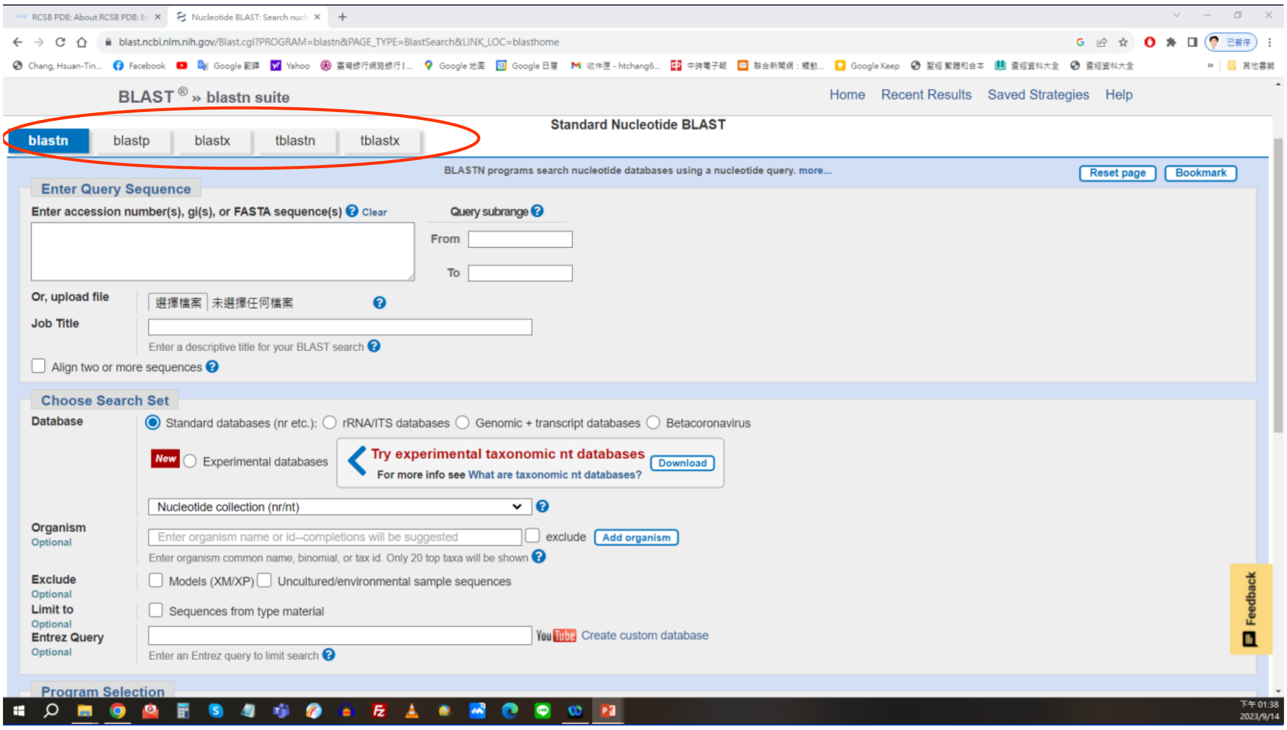
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS
ClusteredNR database on BLAST+
The ClusteredNR database is now available for BLAST+
Thu, 24 Aug 2023
More BLAST news...

Web BLAST

BLAST Genomes

Enter organism common name, scientific name, or tax id | Search



Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange: From / To

Or, upload file: 選擇檔案 未選擇任何檔案

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

New ○ Experimental databases — Try experimental taxonomic nt databases. For more info see What are taxonomic nt databases? Download

Nucleotide collection (nr/nt)

Organism (Optional): Enter organism name or id--completions will be suggested ☐ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

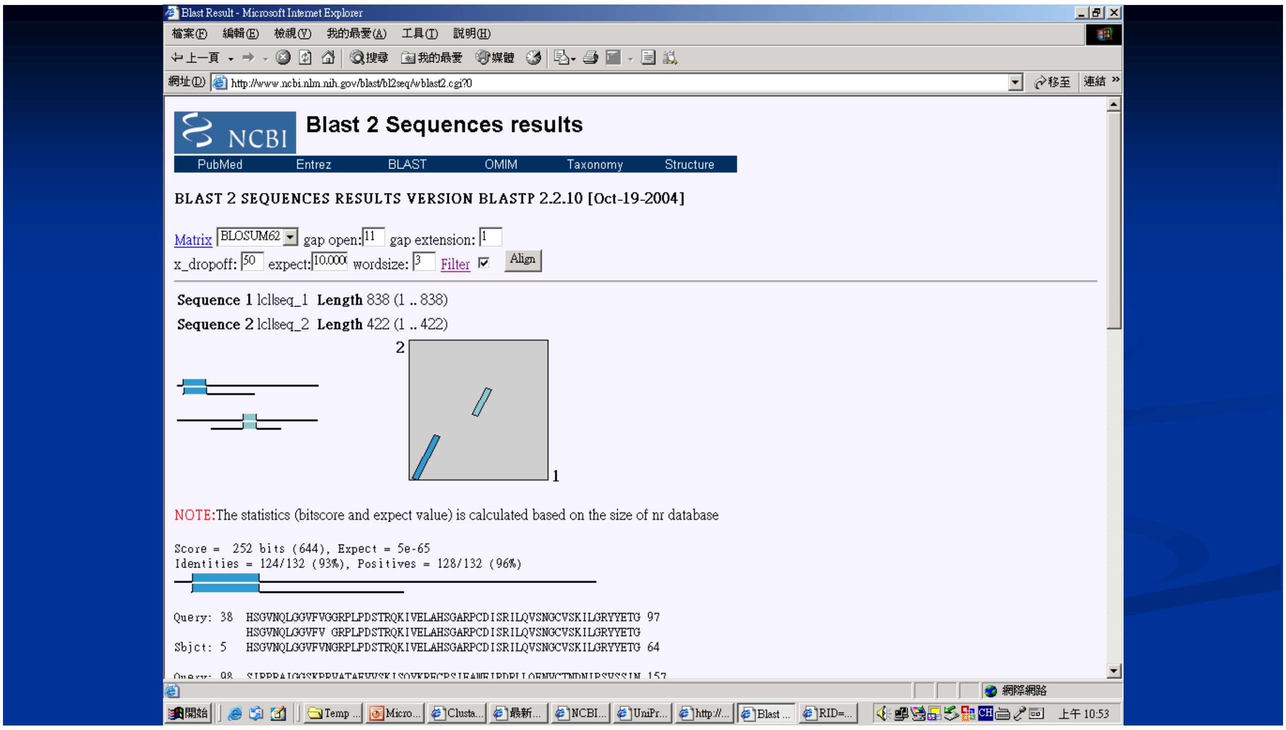
Exclude (Optional): ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to (Optional): ☐ Sequences from type material

Entrez Query (Optional): Create custom database

Enter an Entrez query to limit search

Program Selection



Blast 2 Sequences results

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.10 [Oct-19-2004]

Matrix: BLOSUM62 gap open: 11 gap extension: 1

x_dropoff: 50 expect: 10.000 wordsize: 3 Filter ☑

Sequence 1 lcl|seq_1 **Length** 838 (1 .. 838)

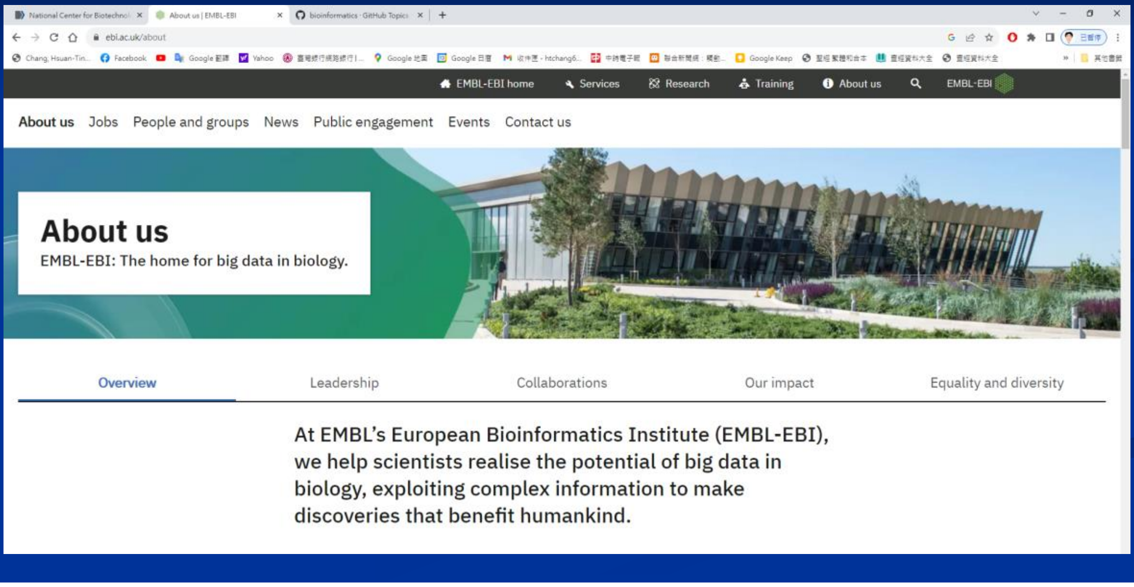
Sequence 2 lcl|seq_2 **Length** 422 (1 .. 422)

NOTE:The statistics (bitscore and expect value) is calculated based on the size of nr database

```
Score =  252 bits (644), Expect = 5e-65
Identities = 124/132 (93%), Positives = 128/132 (96%)

Query: 38   HSGVNQLGGVFVGGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETG 97
             HSGVNQLGGVFV GRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETG
Sbjct: 5    HSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETG 64
```

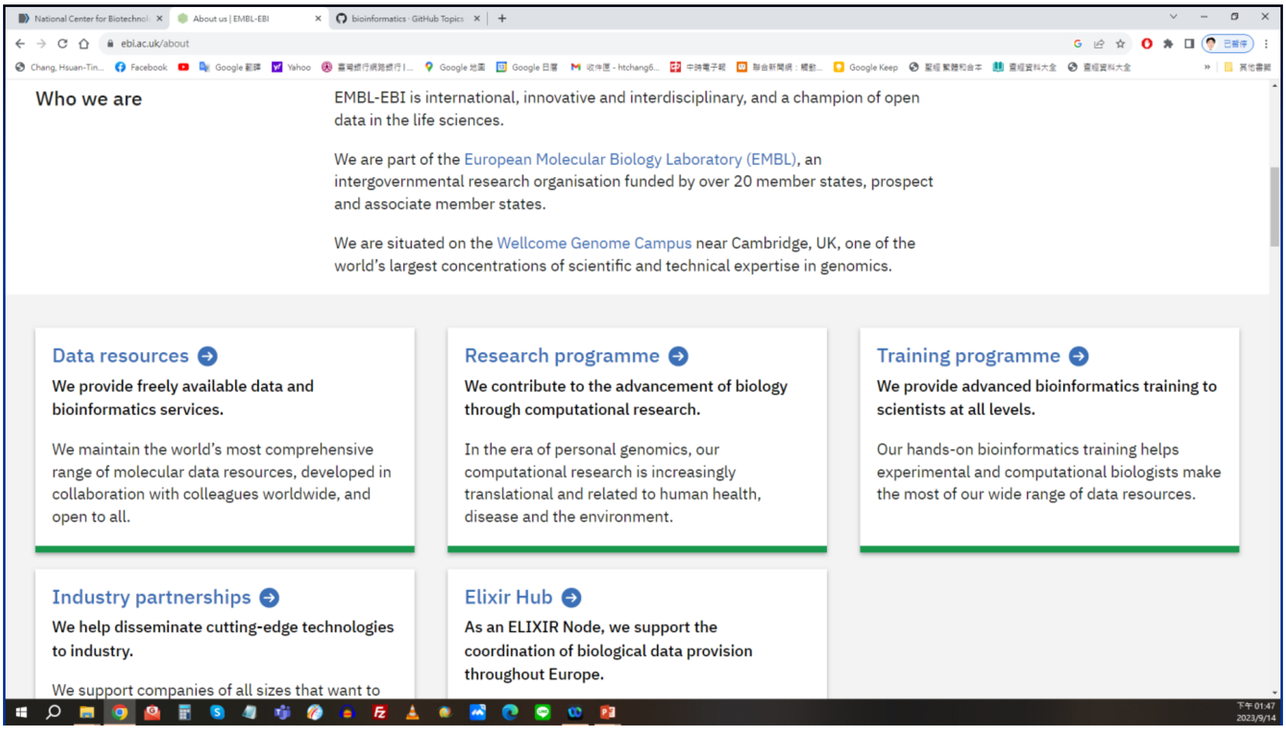
European Bioinformatics Institute



About us

EMBL-EBI: The home for big data in biology.

At EMBL's European Bioinformatics Institute (EMBL-EBI), we help scientists realise the potential of big data in biology, exploiting complex information to make discoveries that benefit humankind.



Who we are

EMBL-EBI is international, innovative and interdisciplinary, and a champion of open data in the life sciences.

We are part of the European Molecular Biology Laboratory (EMBL), an intergovernmental research organisation funded by over 20 member states, prospect and associate member states.

We are situated on the Wellcome Genome Campus near Cambridge, UK, one of the world's largest concentrations of scientific and technical expertise in genomics.

Data resources

We provide freely available data and bioinformatics services.

We maintain the world's most comprehensive range of molecular data resources, developed in collaboration with colleagues worldwide, and open to all.

Research programme

We contribute to the advancement of biology through computational research.

In the era of personal genomics, our computational research is increasingly translational and related to human health, disease and the environment.

Training programme

We provide advanced bioinformatics training to scientists at all levels.

Our hands-on bioinformatics training helps experimental and computational biologists make the most of our wide range of data resources.

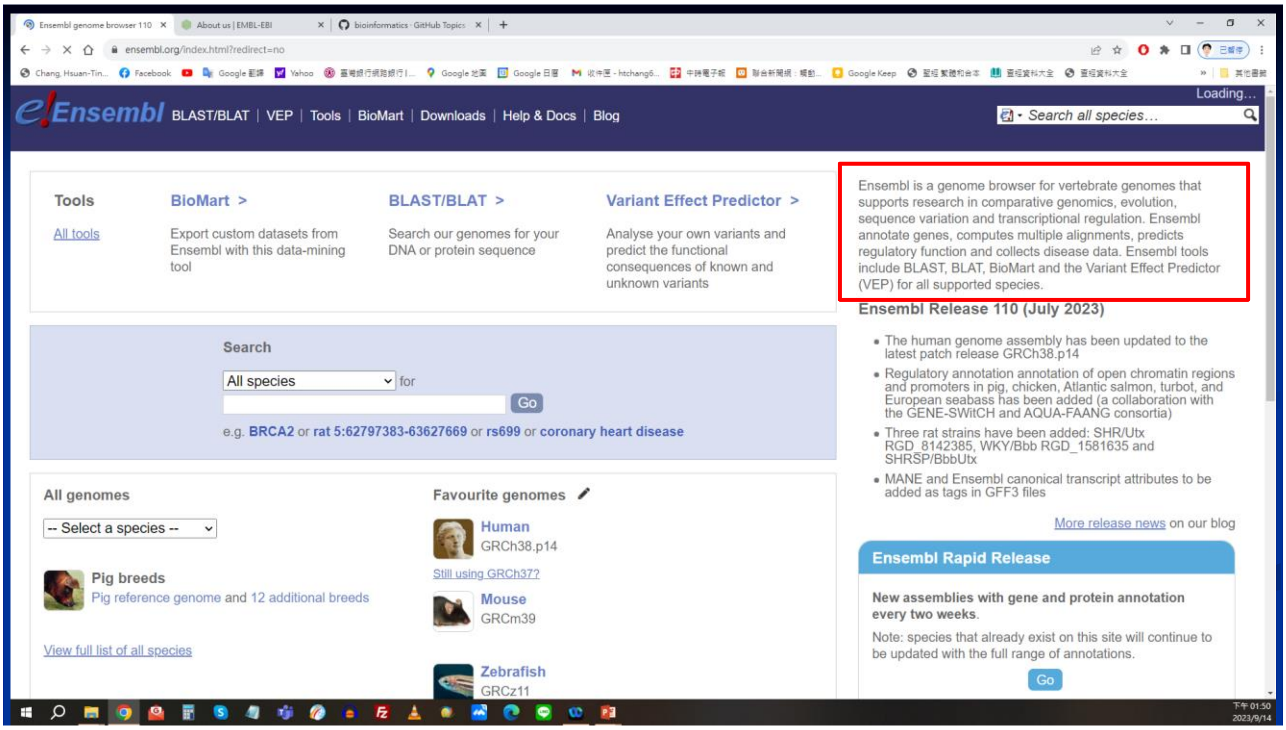
Industry partnerships

We help disseminate cutting-edge technologies to industry.

We support companies of all sizes that want to

Elixir Hub

As an ELIXIR Node, we support the coordination of biological data provision throughout Europe.



Tools

All tools

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 110 (July 2023)

- The human genome assembly has been updated to the latest patch release GRCh38.p14
- Regulatory annotation annotation of open chromatin regions and promoters in pig, chicken, Atlantic salmon, turbot, and European seabass has been added (a collaboration with the GENE-SWitCH and AQUA-FAANG consortia)
- Three rat strains have been added: SHR/Utx RGD_8142385, WKY/Bbb RGD_1581635 and SHRSP/BbbUtx
- MANE and Ensembl canonical transcript attributes to be added as tags in GFF3 files

More release news on our blog

Search

All species for

Go

e.g. **BRCA2** or **rat 5:62797383-63627669** or **rs699** or **coronary heart disease**

All genomes

-- Select a species --

Pig breeds

Pig reference genome and 12 additional breeds

View full list of all species

Favourite genomes

Human
GRCh38.p14
Still using GRCh37?

Mouse
GRCm39

Zebrafish
GRCz11

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

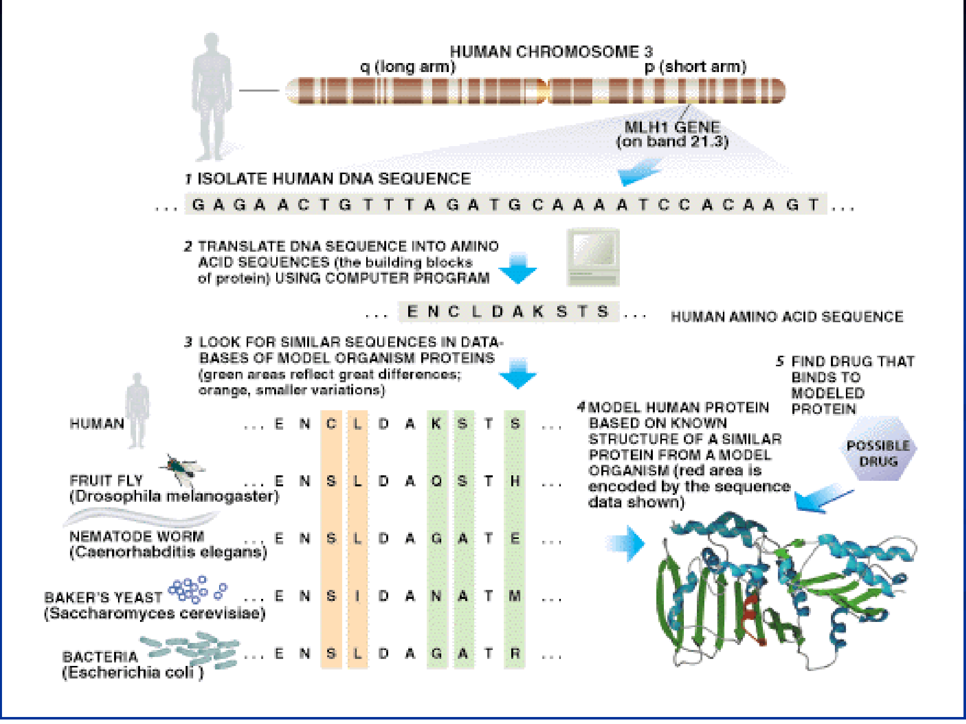
Go

- Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

- Ensembl 是一款脊椎動物基因組瀏覽器，支持比較基因組學、進化、序列變異和轉錄調控方面的研究。 Ensembl 註釋基因、計算多重比對、預測調控功能並收集疾病數據。 Ensembl 工具包括適用於所有受支持物種的 BLAST、BLAT、BioMart 和變異效應預測器 (VEP)。

Applications

- 遺傳學（動植物品種改良）Genetics (animal and plant breed improvement)
- 分子生物學（研究生命現象）Molecular biology (study of life phenomena)
- 臨床醫學（疾病治療）Clinical Medicine (Disease Treatment)
- 生物技術（環境保護與物種複製）Biotechnology (environmental protection and species replication)
- 食品藥物科學（食物改良與藥品設計）Food and Drug Science (Food Improvement and Drug Design)
- 法醫鑑識科學（DNA比對）Food and Drug Science (Food Improvement and Drug Design)
- 考古學（物種演化）Archeology (Evolution of Species)



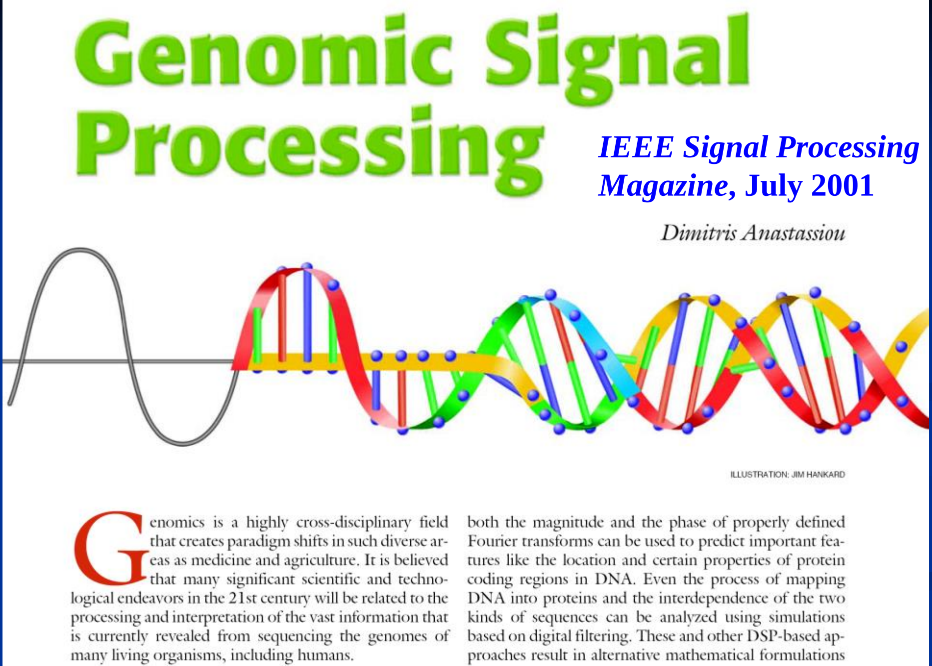
1 ISOLATE HUMAN DNA SEQUENCE

2 TRANSLATE DNA SEQUENCE INTO AMINO ACID SEQUENCES (the building blocks of protein) USING COMPUTER PROGRAM

3 LOOK FOR SIMILAR SEQUENCES IN DATABASES OF MODEL ORGANISM PROTEINS (green areas reflect great differences; orange, smaller variations)

4 MODEL HUMAN PROTEIN BASED ON KNOWN STRUCTURE OF A SIMILAR PROTEIN FROM A MODEL ORGANISM (red area is encoded by the sequence data shown)

5 FIND DRUG THAT BINDS TO MODELED PROTEIN



Genomic Signal Processing

IEEE Signal Processing Magazine, July 2001

Dimitris Anastassiou

ILLUSTRATION: JIM HANKARD

Genomics is a highly cross-disciplinary field that creates paradigm shifts in such diverse areas as medicine and agriculture. It is believed that many significant scientific and technological endeavors in the 21st century will be related to the processing and interpretation of the vast information that is currently revealed from sequencing the genomes of many living organisms, including humans.

both the magnitude and the phase of properly defined Fourier transforms can be used to predict important features like the location and certain properties of protein coding regions in DNA. Even the process of mapping DNA into proteins and the interdependence of the two kinds of sequences can be analyzed using simulations based on digital filtering. These and other DSP-based approaches result in alternative mathematical formulations

---

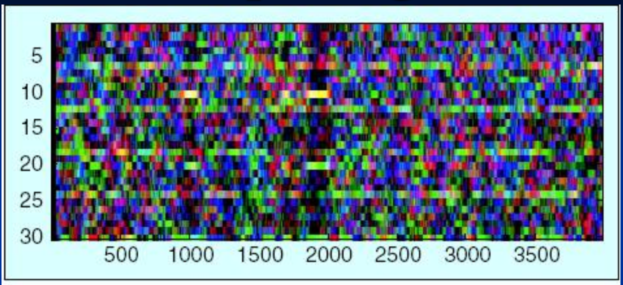
The Problem

- Genomic information is digital letters A, T, C and G
- Signal processing deals with numerical sequences, character strings have to be mapped into one or more numerical sequences
- Identification of protein coding regions
- Prediction of whether or not a given DNA segment is a part of a protein coding region
- Prediction of the proper reading frame
- Comparing to traditional methods, signal processing methods are much quicker, and can be even more accurate in some cases.

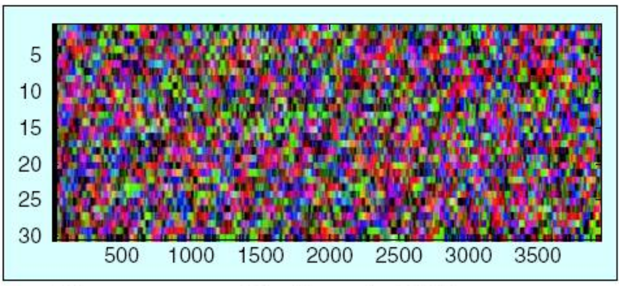
Genomic Signal Processing

- Digital Signal Processing Techniques
  - Fourier transform - spectrogram
  - Wavelet transform
- Visualization & Feature Selection
  - Genomic & protein sequences
  - Sequence representation
  - Database retrieval

DNA Spectrograms



▲ *7. Color spectrogram of a DNA stretch.*



▲ *8. Color spectrogram of "totally random" DNA.*

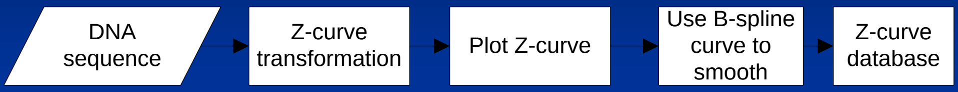
Z curve

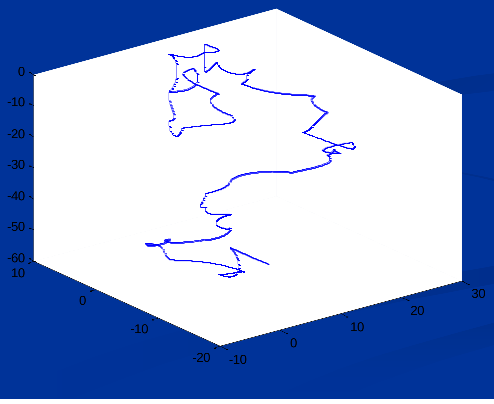
Bioinformatics Vol. 19 no. 5 2003
Pages 593-599
© 2003 Oxford University Press

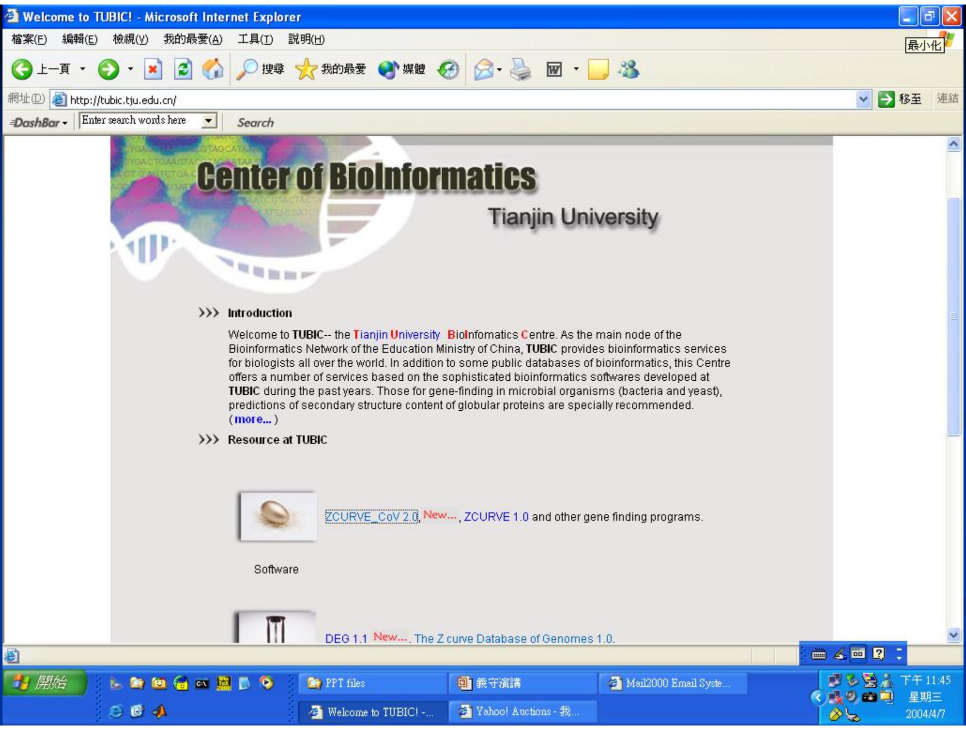
The Z curve database: a graphic representation of genome sequences
Chun-Ting Zhang 1,*, Ren Zhang 2 and Hong-Yu Ou 1
1 Department of Physics, Tianjin University, Tianjin 300072, People's Republic of China
2 Department of Epidemiology and Biostatistics, Tianjin Cancer Institute and Hospital, Tianjin 300060, People's Republic of China

$$\begin{cases} x_n = (A_n + G_n) - (C_n + T_n), \\ y_n = (A_n + C_n) - (G_n + T_n), \quad x_n,\ y_n,\ z_n \in [-N,\ N],\ n = 0,\ 1,\ 2, \ldots,\ N, \\ z_n = (A_n + T_n) - (C_n + G_n), \end{cases}$$

Z-curve flowchart


DNA sequence
Z-curve transformation
Plot Z-curve
Use B-spline curve to smooth
Z-curve database





Center of BioInformatics

Tianjin University

>>> Introduction

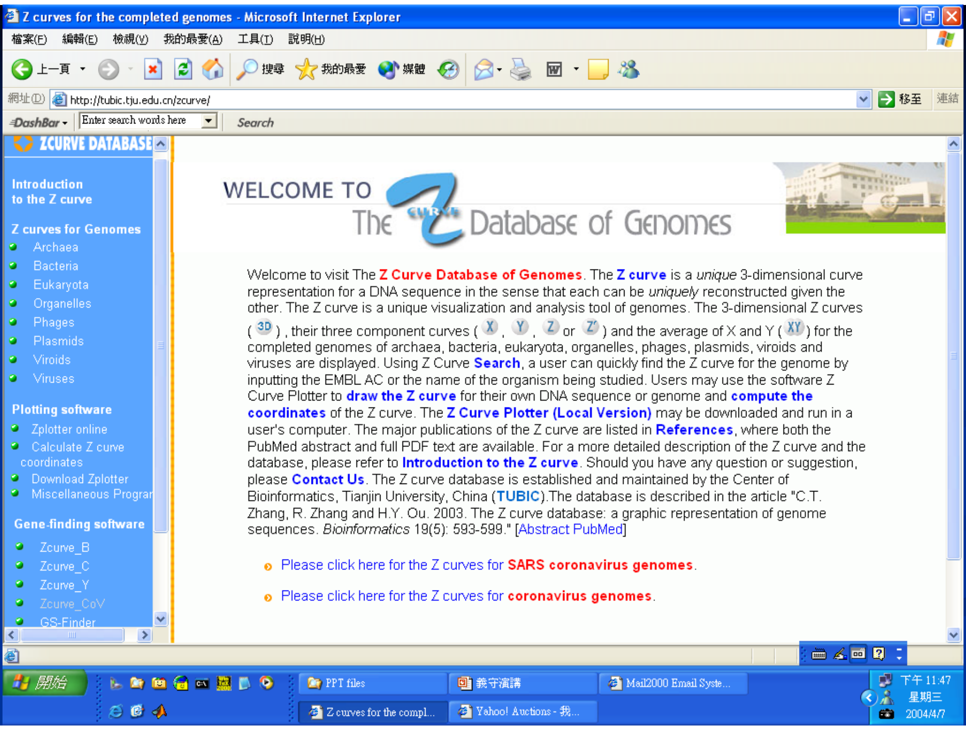
Welcome to **TUBIC**-- the Tianjin University BioInfomatics Centre. As the main node of the Bioinformatics Network of the Education Ministry of China, **TUBIC** provides bioinformatics services for biologists all over the world. In addition to some public databases of bioinformatics, this Centre offers a number of services based on the sophisticated bioinformatics softwares developed at **TUBIC** during the past years. Those for gene-finding in microbial organisms (bacteria and yeast), predictions of secondary structure content of globular proteins are specially recommended. (**more...**)

>>> Resource at TUBIC

ZCURVE_CoV 2.0, New..., ZCURVE 1.0 and other gene finding programs.

Software

DEG 1.1 New..., The Z curve Database of Genomes 1.0.



Introduction to the Z curve

Z curves for Genomes

- Archaea
- Bacteria
- Eukaryota
- Organelles
- Phages
- Plasmids
- Viroids
- Viruses

Plotting software

- Zplotter online
- Calculate Z curve coordinates
- Download Zplotter
- Miscellaneous Prograr

Gene-finding software

- Zcurve_B
- Zcurve_C
- Zcurve_Y
- Zcurve_CoV
- GS-Finder

WELCOME TO

The Z CURVE Database of Genomes

Welcome to visit The **Z Curve Database of Genomes**. The **Z curve** is a *unique* 3-dimensional curve representation for a DNA sequence in the sense that each can be *uniquely* reconstructed given the other. The Z curve is a unique visualization and analysis tool of genomes. The 3-dimensional Z curves (3D) , their three component curves (X , Y , Z or Z') and the average of X and Y (XY) for the completed genomes of archaea, bacteria, eukaryota, organelles, phages, plasmids, viroids and viruses are displayed. Using Z Curve **Search**, a user can quickly find the Z curve for the genome by inputting the EMBL AC or the name of the organism being studied. Users may use the software Z Curve Plotter to **draw the Z curve** for their own DNA sequence or genome and **compute the coordinates** of the Z curve. The **Z Curve Plotter (Local Version)** may be downloaded and run in a user's computer. The major publications of the Z curve are listed in **References**, where both the PubMed abstract and full PDF text are available. For a more detailed description of the Z curve and the database, please refer to **Introduction to the Z curve**. Should you have any question or suggestion, please **Contact Us**. The Z curve database is established and maintained by the Center of Bioinformatics, Tianjin University, China (**TUBIC**).The database is described in the article "C.T. Zhang, R. Zhang and H.Y. Ou. 2003. The Z curve database: a graphic representation of genome sequences. *Bioinformatics* 19(5): 593-599." [Abstract PubMed]

- Please click here for the Z curves for **SARS coronavirus genomes**.
- Please click here for the Z curves for **coronavirus genomes**.

利用三維軌跡作為DNA序列的視覺化與比對

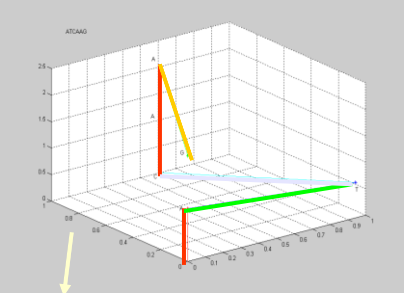
Visualization and Comparison of DNA Sequences by Use of Three-Dimensional Trajectories

Presented in 2003 Asia Pacific Bioinformatics Conference (APBC2003)
Adelaide Australia, Feb 2003

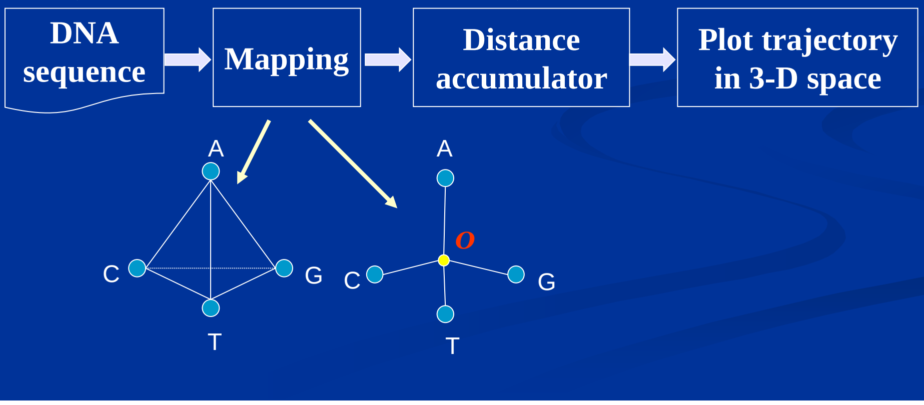
Three-dimensional trajectory (TDT)

- A global view visualization tool
- Presentation of ***pseudo shape*** of DNA sequences
- Low computational overhead
- Easy to find major similarity/difference among DNA sequences
- Use DSP algorithms for further processing

Algorithm



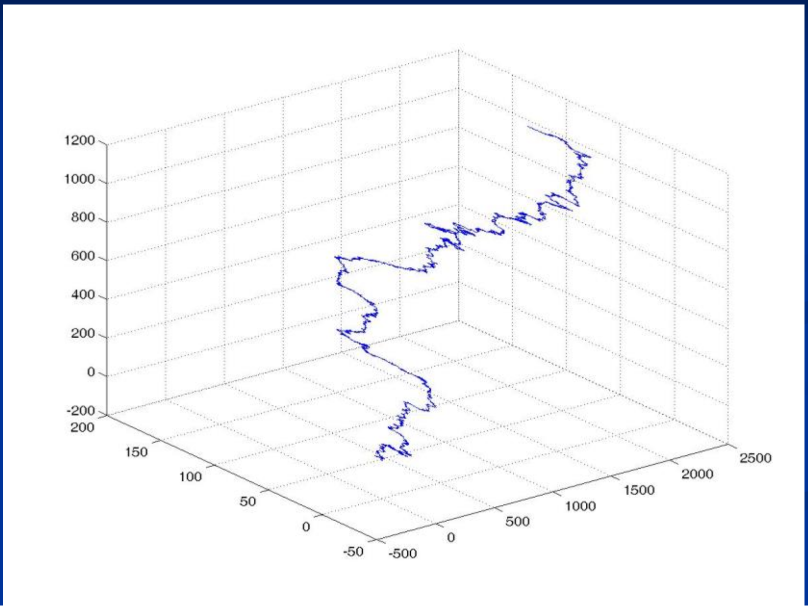
TTGCAAAGGTGACACTTGTAT
TTGCAACTGTGATGTTTG
GG….

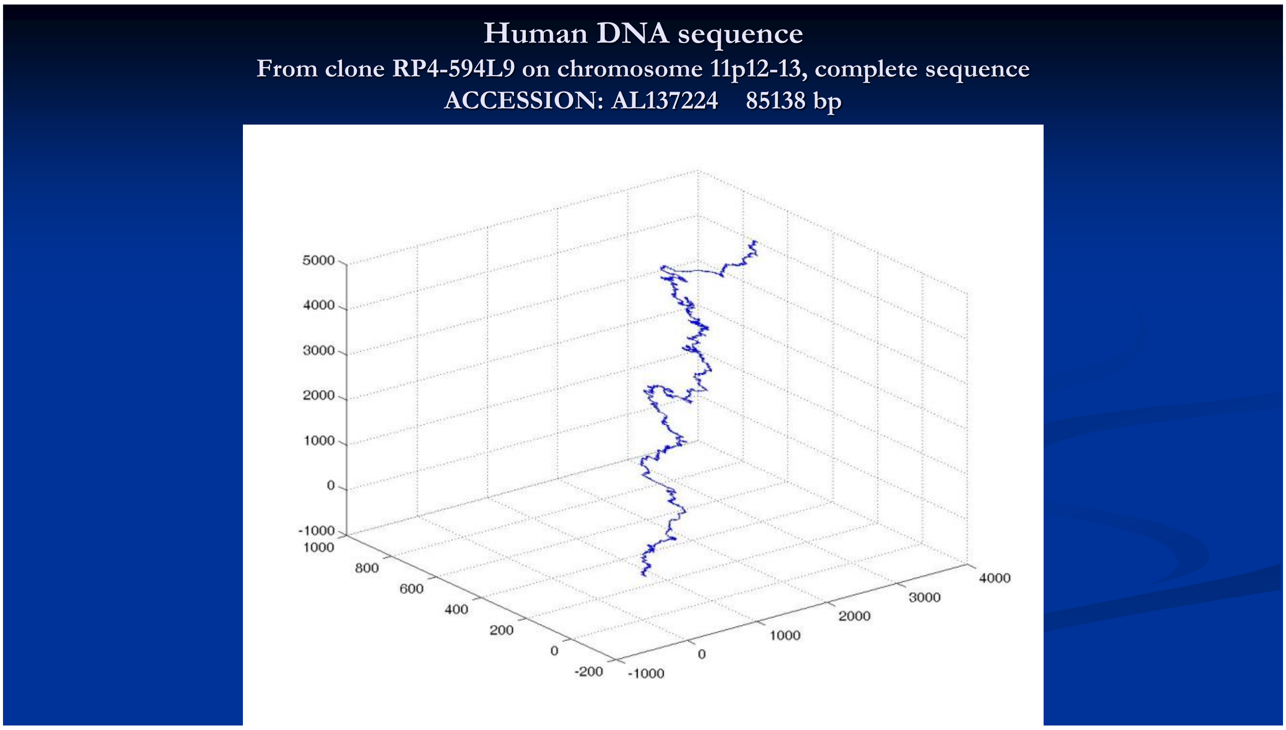


DNA sequence → **Mapping** → **Distance accumulator** → **Plot trajectory in 3-D space**

Human DNA sequence

Homo sapiens chromosome 1 clone RP11-20L9, complete sequence.
ACCESSION: AC095037, 15000 bp

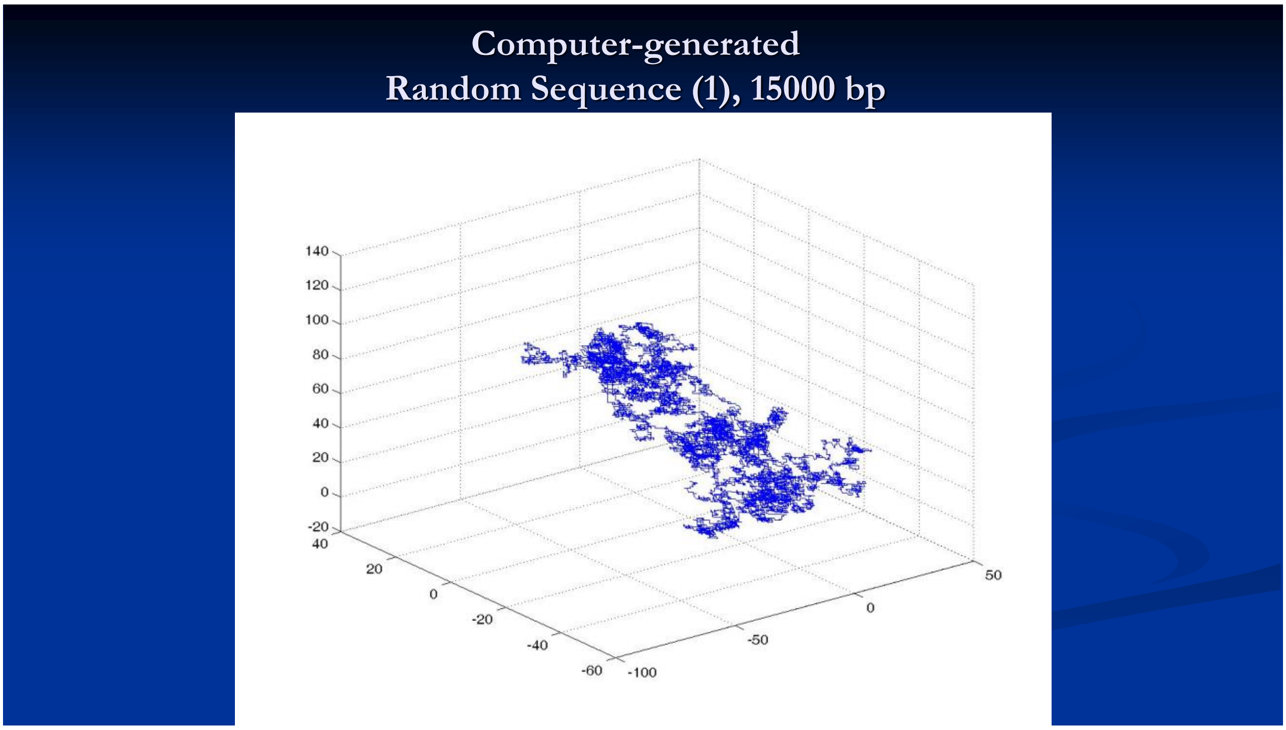




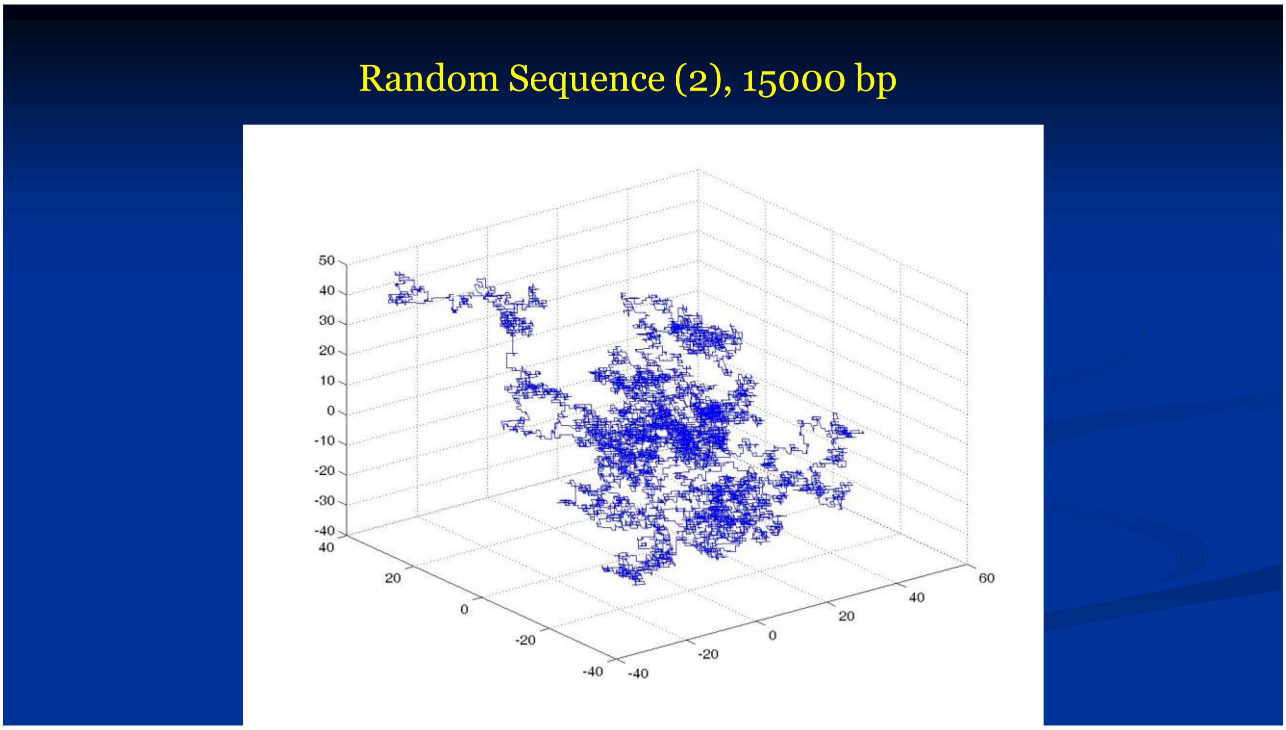
Human DNA sequence

From clone RP4-594L9 on chromosome 11p12-13, complete sequence

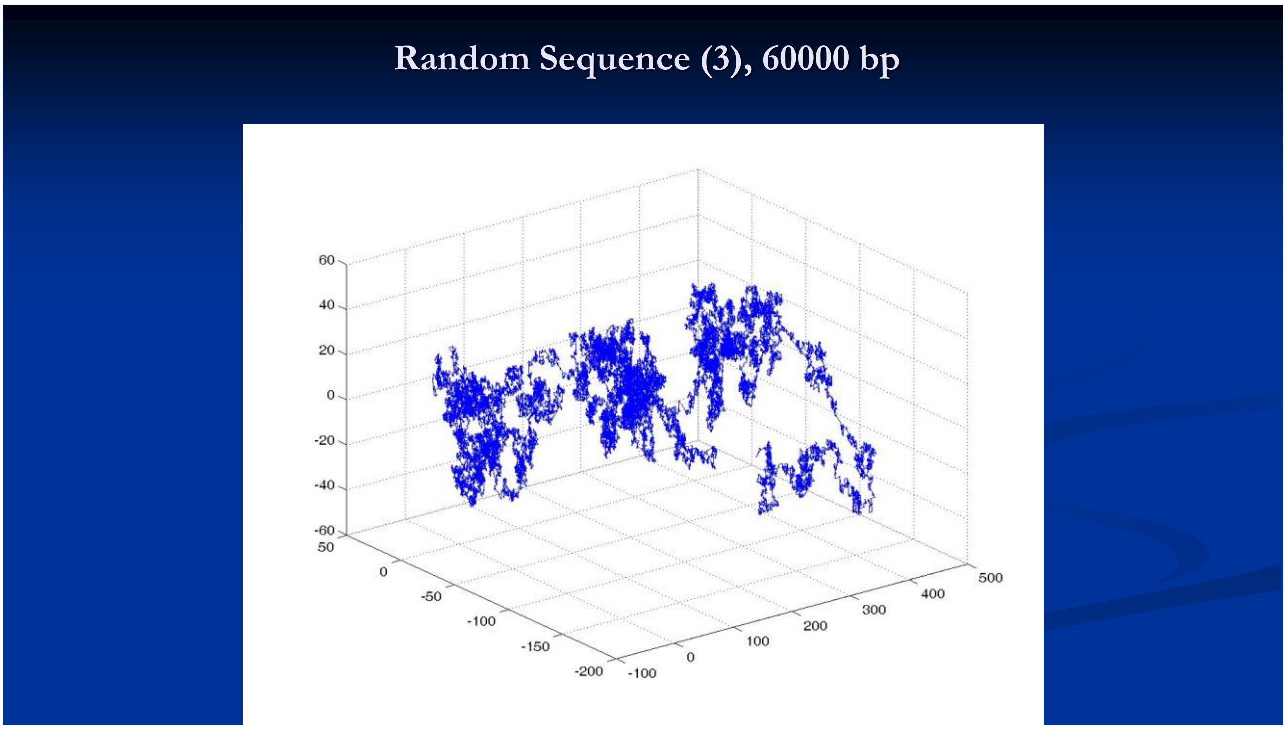
ACCESSION: AL137224   85138 bp



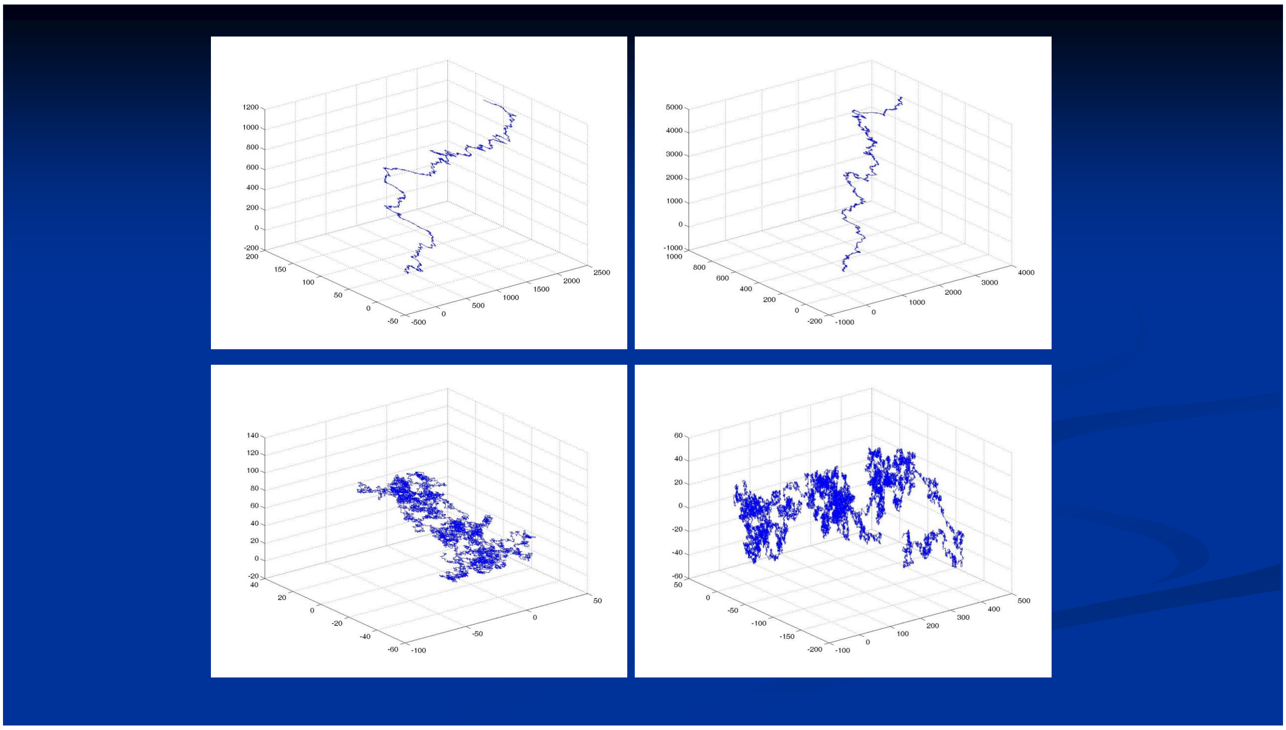
Computer-generated
Random Sequence (1), 15000 bp

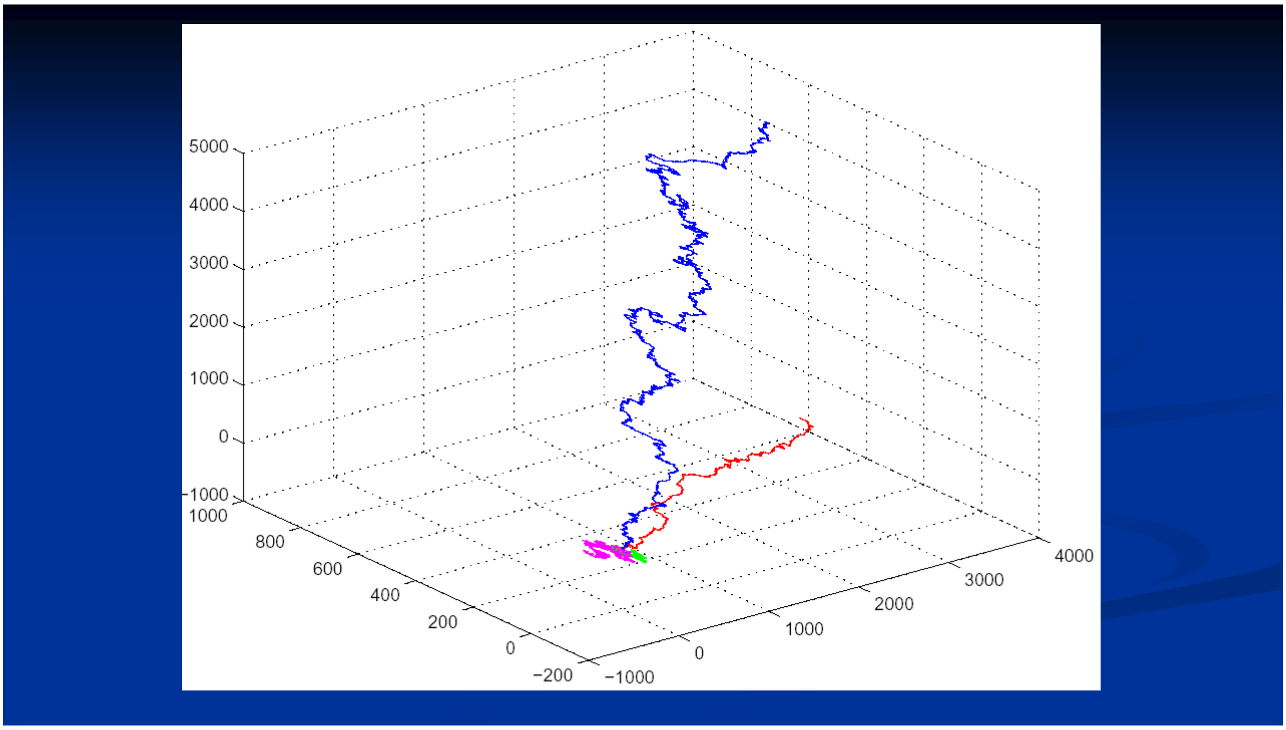


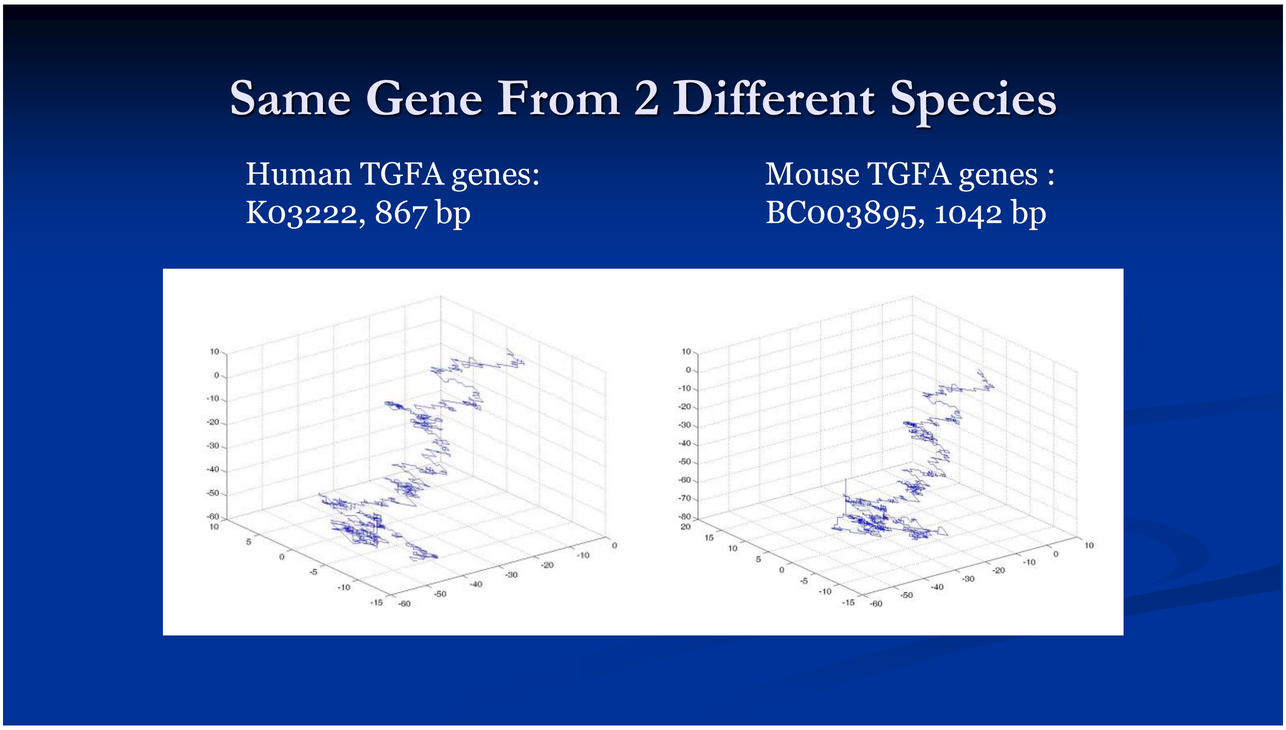
Random Sequence (2), 15000 bp



Random Sequence (3), 60000 bp


1200
1000
800
600
400
200
0
-200
5000
4000
3000
2000
1000
0
-1000
140
120
100
80
60
40
20
0
-20
60
40
20
0
-20
-40
-60

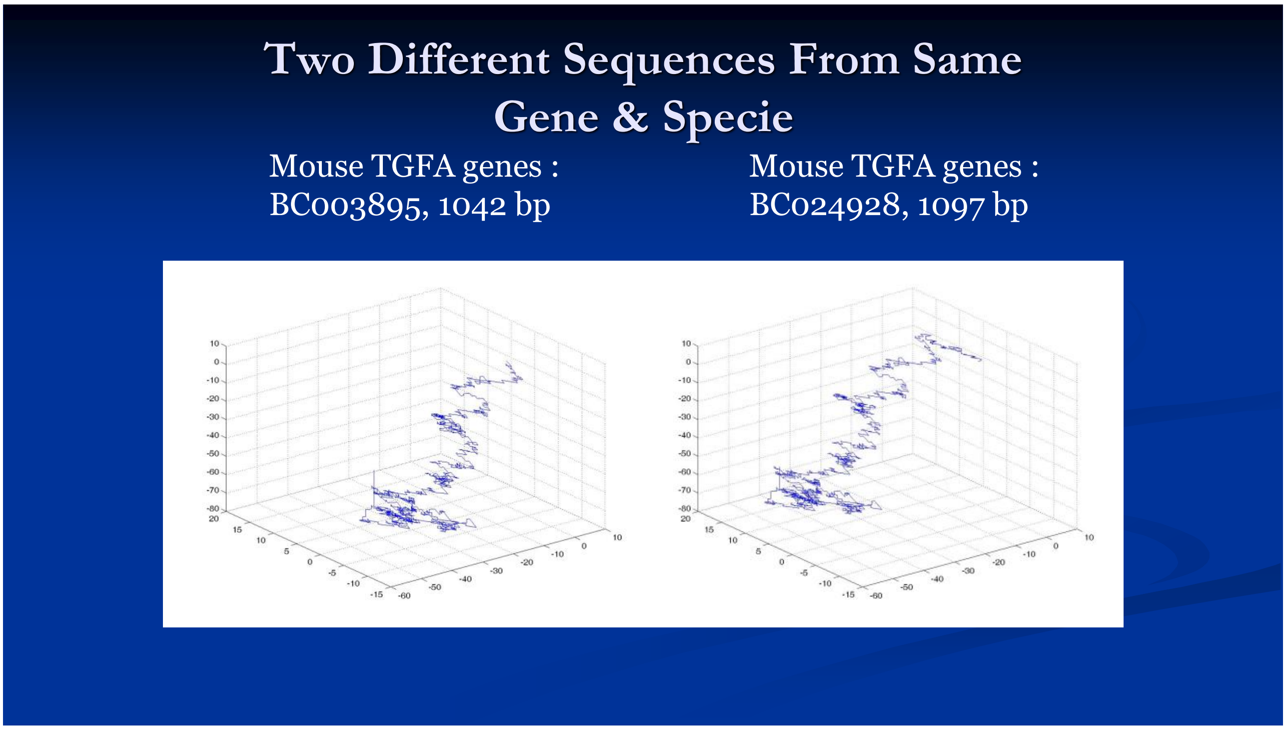

5000
4000
3000
2000
1000
0
−1000
1000
800
600
400
200
0
−200
−1000
0
1000
2000
3000
4000



Same Gene From 2 Different Species

Human TGFA genes:
K03222, 867 bp

Mouse TGFA genes :
BC003895, 1042 bp



Two Different Sequences From Same Gene & Specie

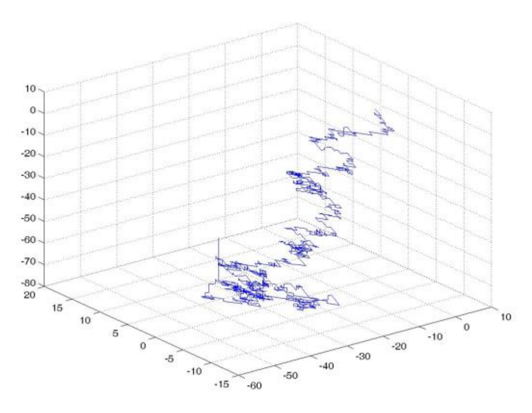
Mouse TGFA genes :
BC003895, 1042 bp

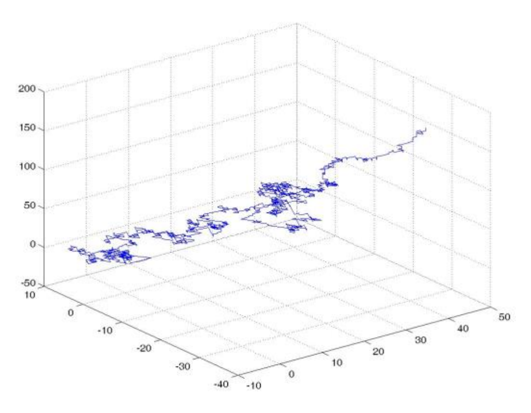
Mouse TGFA genes :
BC024928, 1097 bp

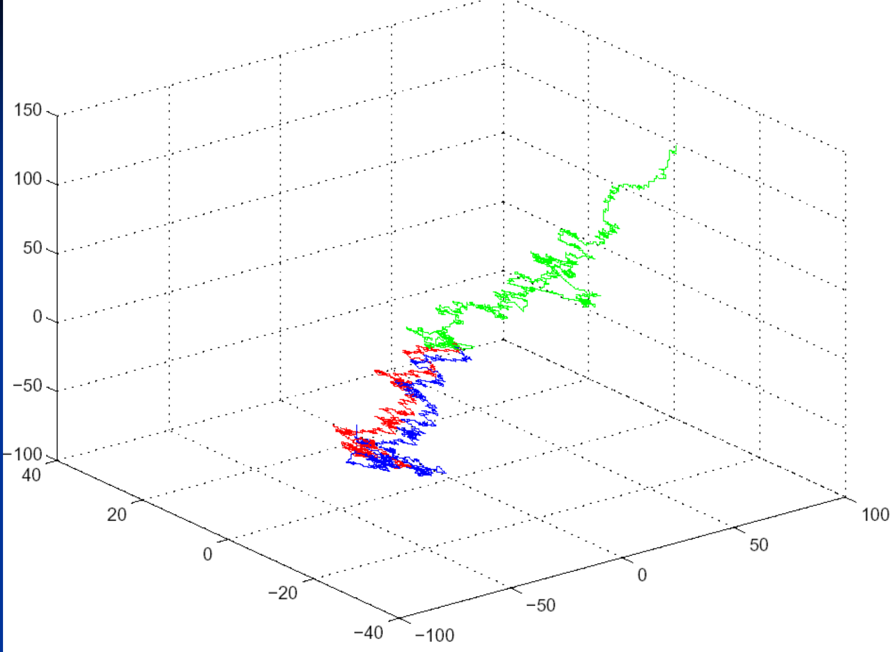
Different Sequences From One Specie

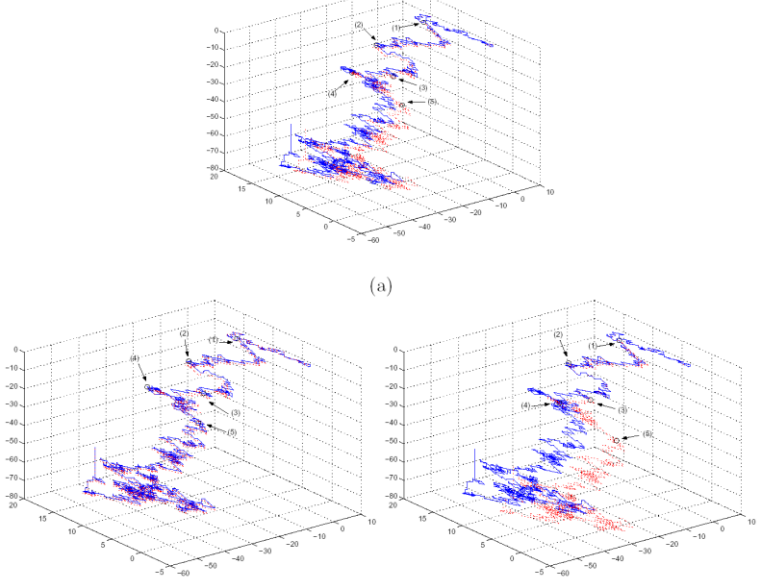
Mouse TGFA genes:
BC003895, 1042 bp

Mouse DHFR genes:
L26316, 1499 bp







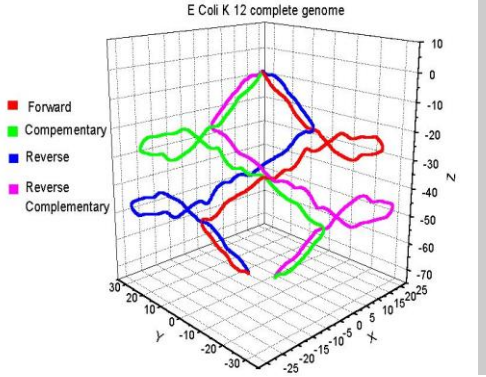


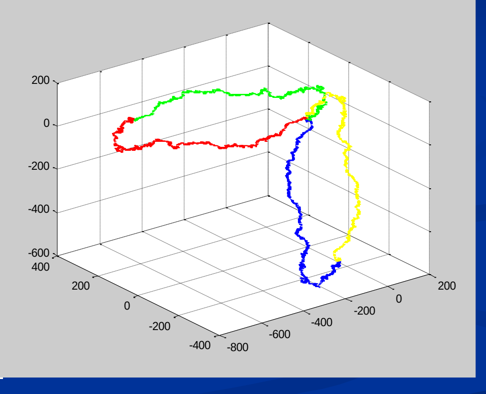
(a)

(b) (c)

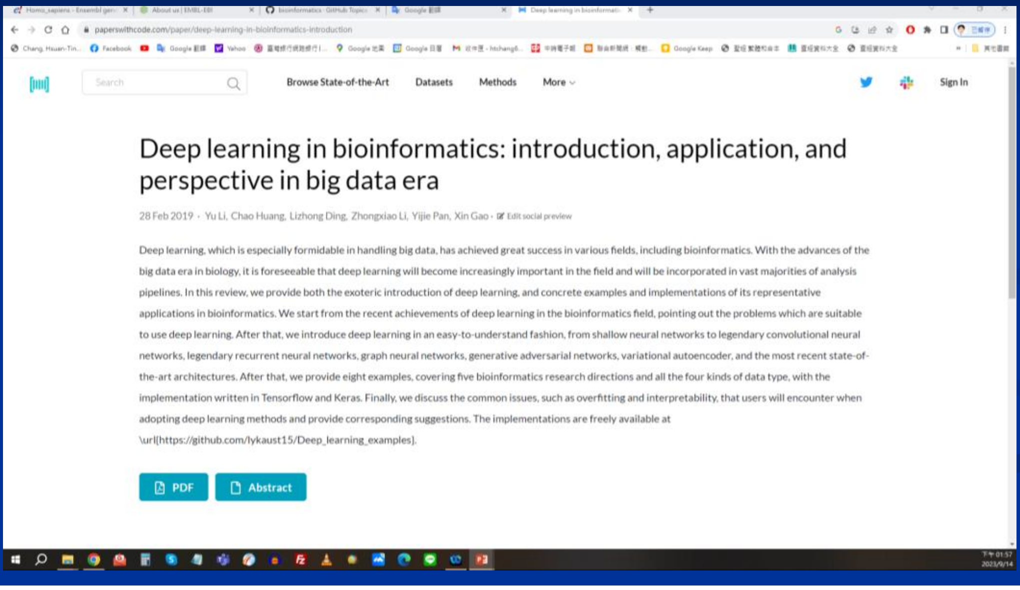
Figure 7: The TDTs for the mouse TGFA gene BC031432, 1109 bp, with (a) 5 bps deletion, (b) 5 bps insertion, and (c) 5 bps substitution.

Z curve v.s. TDT





Deep Learning Techniques on Bioinformatics

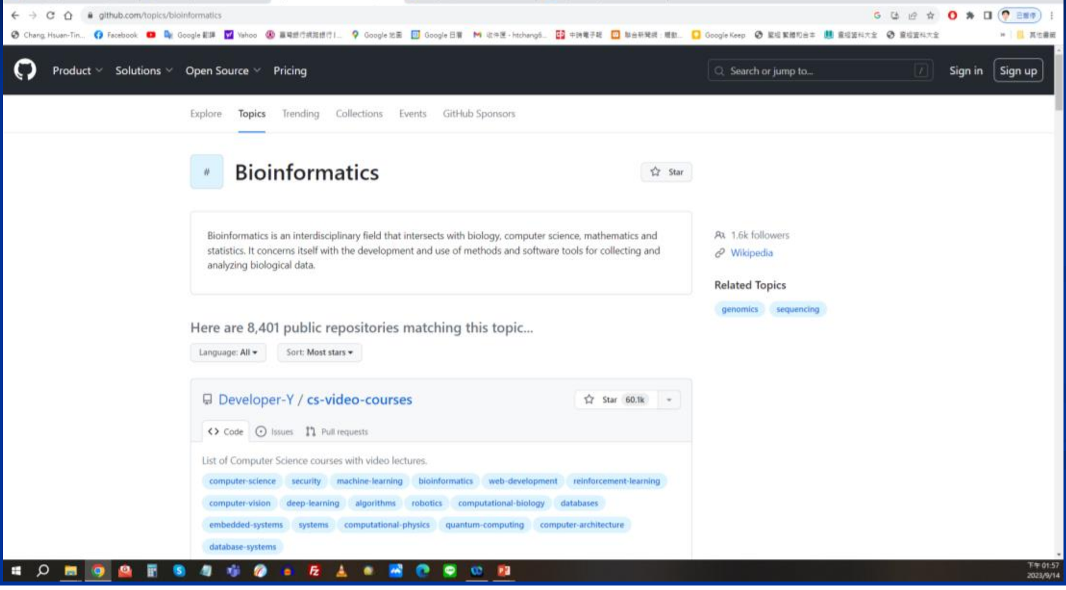


Deep learning in bioinformatics: introduction, application, and perspective in big data era

28 Feb 2019 · Yu Li, Chao Huang, Lizhong Ding, Zhongxiao Li, Yijie Pan, Xin Gao

Deep learning, which is especially formidable in handling big data, has achieved great success in various fields, including bioinformatics. With the advances of the big data era in biology, it is foreseeable that deep learning will become increasingly important in the field and will be incorporated in vast majorities of analysis pipelines. In this review, we provide both the exoteric introduction of deep learning, and concrete examples and implementations of its representative applications in bioinformatics. We start from the recent achievements of deep learning in the bioinformatics field, pointing out the problems which are suitable to use deep learning. After that, we introduce deep learning in an easy-to-understand fashion, from shallow neural networks to legendary convolutional neural networks, legendary recurrent neural networks, graph neural networks, generative adversarial networks, variational autoencoder, and the most recent state-of-the-art architectures. After that, we provide eight examples, covering five bioinformatics research directions and all the four kinds of data type, with the implementation written in Tensorflow and Keras. Finally, we discuss the common issues, such as overfitting and interpretability, that users will encounter when adopting deep learning methods and provide corresponding suggestions. The implementations are freely available at \url{https://github.com/lykaust15/Deep_learning_examples}.

GitHub topics related



Bioinformatics

Bioinformatics is an interdisciplinary field that intersects with biology, computer science, mathematics and statistics. It concerns itself with the development and use of methods and software tools for collecting and analyzing biological data.

Here are 8,401 public repositories matching this topic...

List of Computer Science courses with video lectures.

Challenges & Future Work

- Genomic signal processing opens a new signal processing frontier
- Sequence analysis: symbolic or categorical signal, classical signal processing methods are not directly applicable
- Increasingly high dimensionality of genetic data sets and the complexity involved call for fast and high throughput implementations of genomic signal processing algorithms
- Future work: spectral analysis of DNA sequence and data clustering of microarray data. Modify classical signal processing methods, and develop new ones.

What are the major Challenges & Future Work of bioinformatics research?

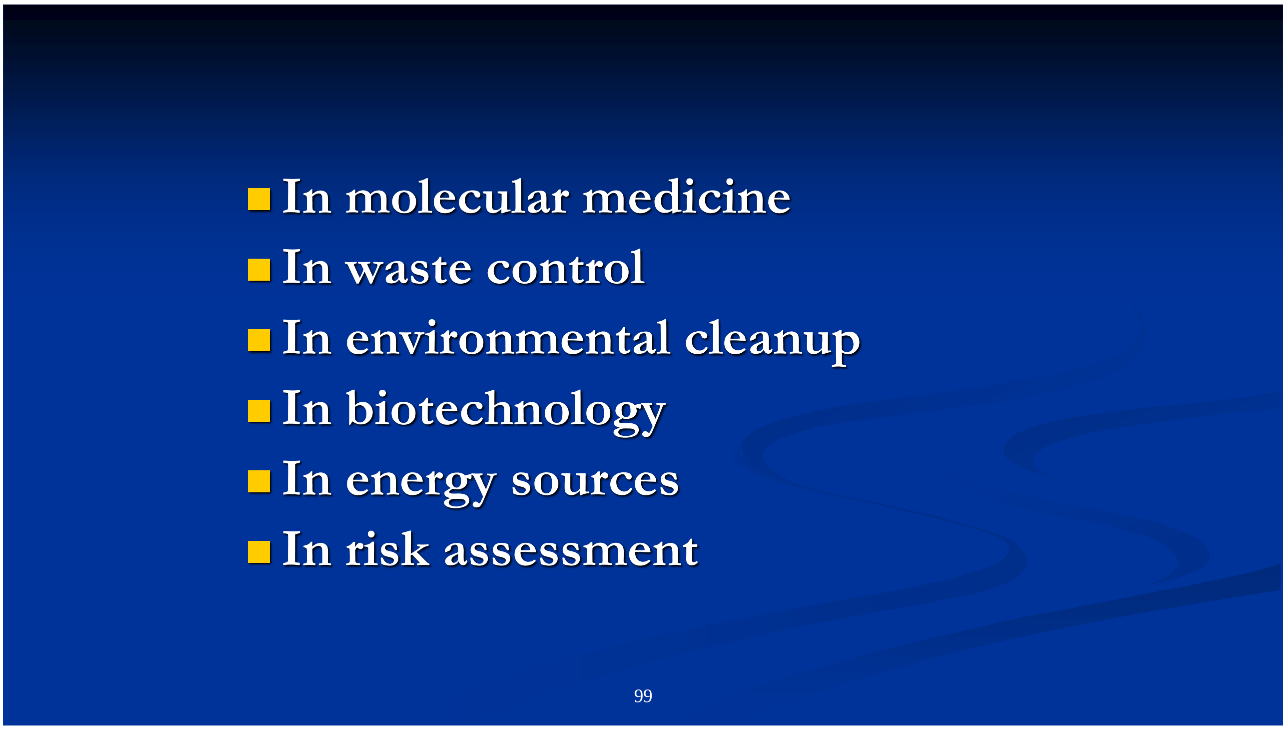
Bioinformatics research faces several challenges and opportunities for future work, including:

1. Data management: As the amount of biological data continues to grow exponentially, managing and storing this data is becoming increasingly challenging. Future work in bioinformatics will focus on developing new tools and techniques for data management.
2. Integration of data: Integrating data from multiple sources is another major challenge in bioinformatics. Future work will focus on developing new methods for integrating and analyzing data from different sources.

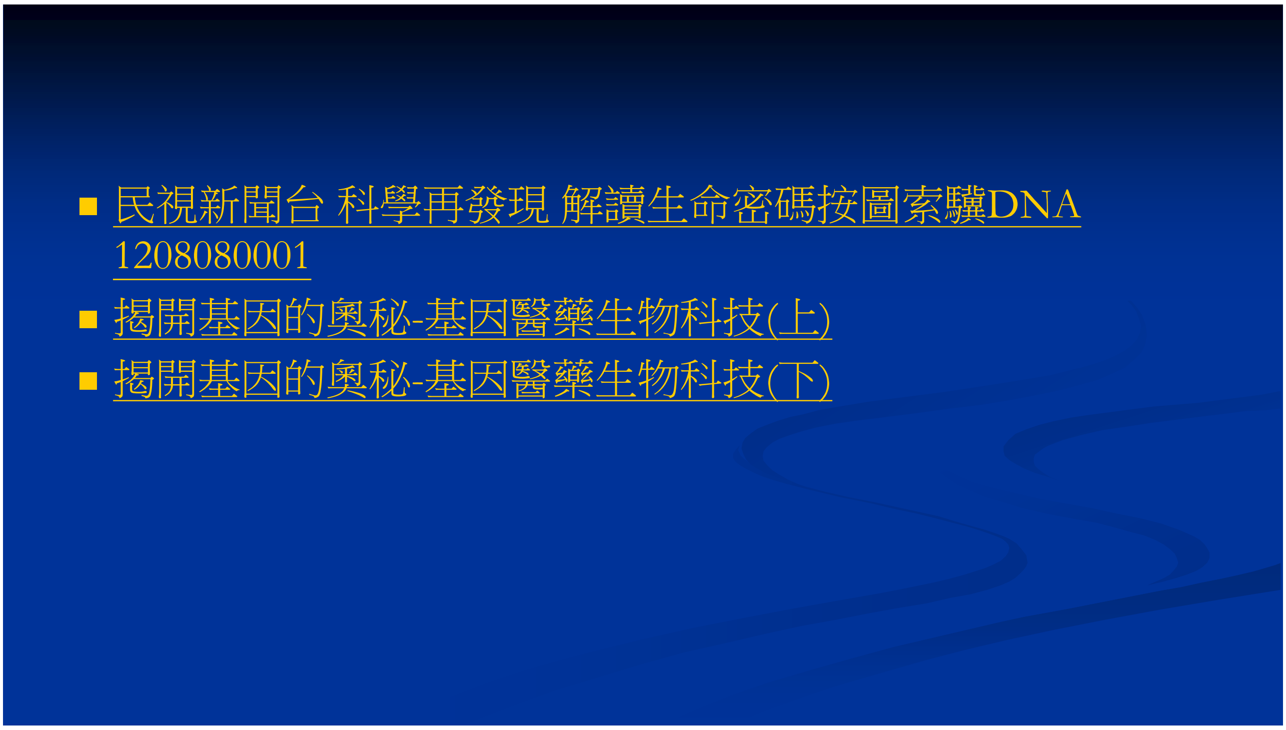
3. Algorithm development: Developing new algorithms for analyzing biological data is a major focus of bioinformatics research. Future work in this area will focus on developing new algorithms that can handle increasingly complex datasets.
4. Interdisciplinary collaboration: Bioinformatics research requires collaboration between researchers from different disciplines, including biology, computer science, and statistics. Future work will focus on fostering interdisciplinary collaboration to advance the field.

5. Ethical considerations: As bioinformatics research becomes more advanced, ethical considerations are becoming increasingly important. Future work will focus on developing ethical guidelines for bioinformatics research.

- Overall, the future of bioinformatics research looks promising, with opportunities for new discoveries and advancements in the field.



- In molecular medicine
- In waste control
- In environmental cleanup
- In biotechnology
- In energy sources
- In risk assessment



- 民視新聞台 科學再發現 解讀生命密碼按圖索驥DNA 1208080001
- 揭開基因的奧秘-基因醫藥生物科技(上)
- 揭開基因的奧秘-基因醫藥生物科技(下)

TED Talks about bioinformatics

- 台灣的生物科技夢：林政和 (Jimmy Lin) at TEDxTaipei 2013
- The Human Genome: Collaboration is the New Competition | Dr. David Haussler | TEDxSantaCruz
- How to interpret the human genome | Alisha Holloway | TEDxClaremontColleges
- Bioinformatics: A way to deciphere DNA and cure life's deadliest diseases | Spencer Hall | TEDxUGA

相關延伸演講

- 當深度學習與生物資訊在個人全基因體註解相遇
- 主講人：陳倩瑜 臺大生物產業機電工程學系教授
- https://www.youtube.com/watch?v=JwTYvtUTiRM

Thank you very much for your attention!