



生物信息学

第4讲 转录组学与调控网络分析（2）

张高川

遗传学与生物信息学系
基础医学与生物科学学院
苏州大学医学部

Email: zhanggaochuan@suda.edu.cn

Tel: 18962111592

WX: [zhanggaochuan770609](#)

QQ: 2257916241



转录组研究方法简介 (2)

>> 高通量技术和方法

**DNA microarray
(DNA chip or gene chip)**

**Serial analysis of gene expression
(SAGE)**

RNA-Seq

EST/cDNA library

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



Introduction to ESTs and cDNA library

1980s

cDNA序列进行大规模测序



Apply knowledge of
microbial functional
capabilities

Protect workers
and the public



Introduction to ESTs and cDNA library

1980s

cDNA序列进行大规模测序



Protect workers
and the public

Apply knowledge of
microbial functional
capabilities

INNOVATIVE APPROACHES
ALONG UNCONVENTIONAL PATHS
U.S. DEPARTMENT OF ENERGY

DNA SEQUENCE DATA
FROM GENOME PROJECTS

carbon

goal
EXPLORE
FUNCTIONS
IN MICROBIAL
COMMUNITIES

COMMUNITY
OF CELLS

goal
IDENTIFY
PROTEIN
MACHINES
DEVELOP
COMPUTATIONAL
CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

WORKING
CELL

Many protein
machines interact
through complex,
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life
processes.

Genes and other
DNA sequences
contain instructions
on how and when
to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



Introduction to ESTs and cDNA library

1980s

cDNA序列进行大规模测序

能发现成千
上万的**新基
因**



Protect workers
and the public

Apply knowledge of
microbial functional
capabilities



Genes
DNA sequences
contain
information
on how cells
use them
to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY
PROTEIN
MACHINES

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

goal
DEVELOP
COMPUTATIONAL
CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

goal
EXPLORE
FUNCTIONS
IN MICROBIAL
COMMUNITIES

COMMUNITY
OF CELLS

WORKING
CELL

Many protein
machines interact
through complex,
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life
processes.

URL: DOEGenomesToLife.org

10/02



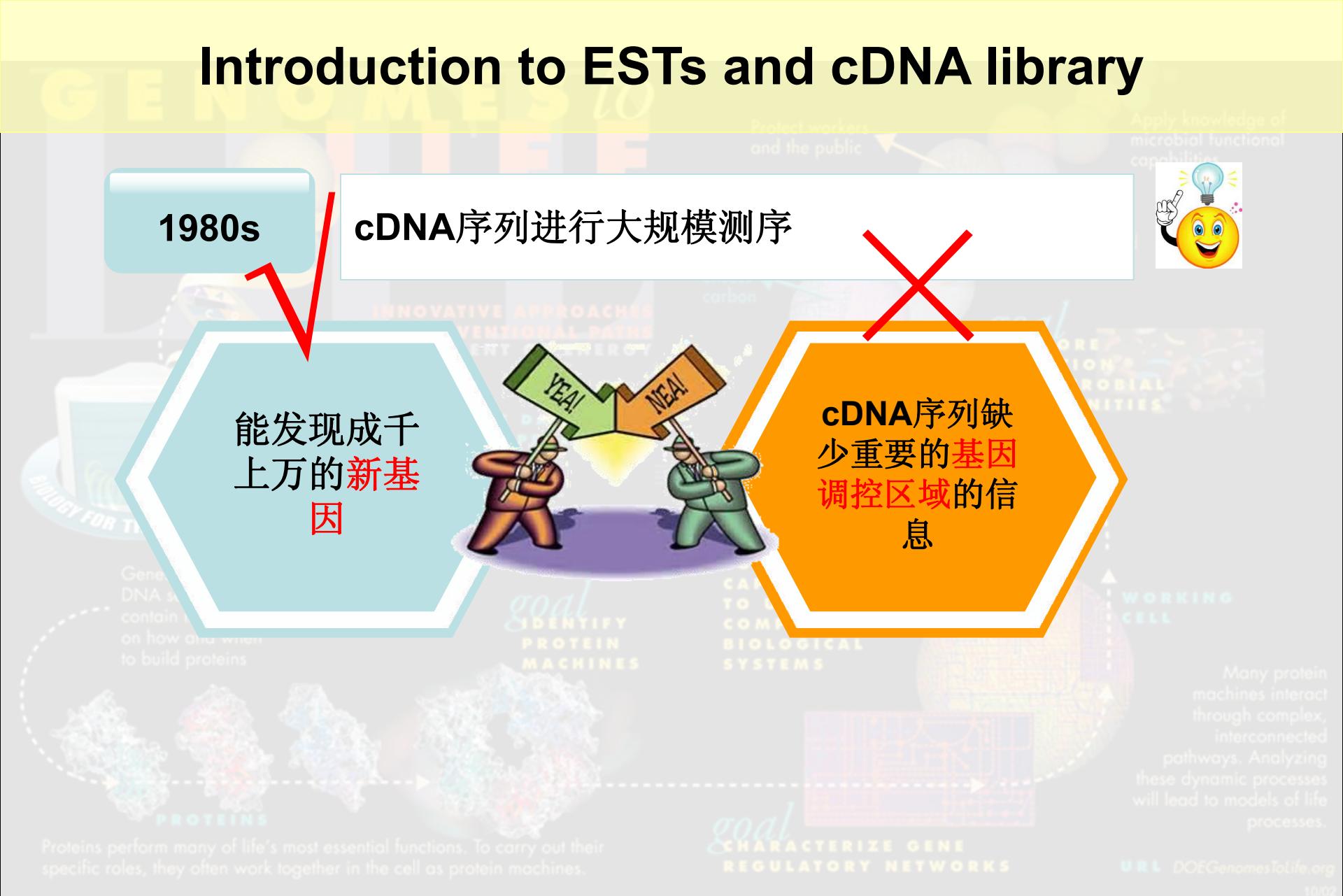
Introduction to ESTs and cDNA library

1980s

cDNA序列进行大规模测序

能发现成千上万的**新基因**

cDNA序列缺少重要的**基因调控区域**的信息



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

URL: DOEGenomesToLife.org

10/02



Introduction to ESTs and cDNA library

1980s

cDNA序列进行大规模测序

能发现成千上万的**新基因**

cDNA序列缺少重要的**基因调控区域**的信息



90年代初, **Graig Venter** 提出了EST的概念, 并测定了609条人脑组织的EST, 宣布了cDNA大规模测序的时代的开始 (**Adams et al., 1991, Science**)。

cDNA library

A cDNA library is a combination of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which together constitute some portion of the transcriptome of the organism.

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Apply knowledge of microbial functional capabilities

Produce and use energy

EXPLORE FUNCTIONS IN MICROBIAL COMMUNITIES

WORKING CELL

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

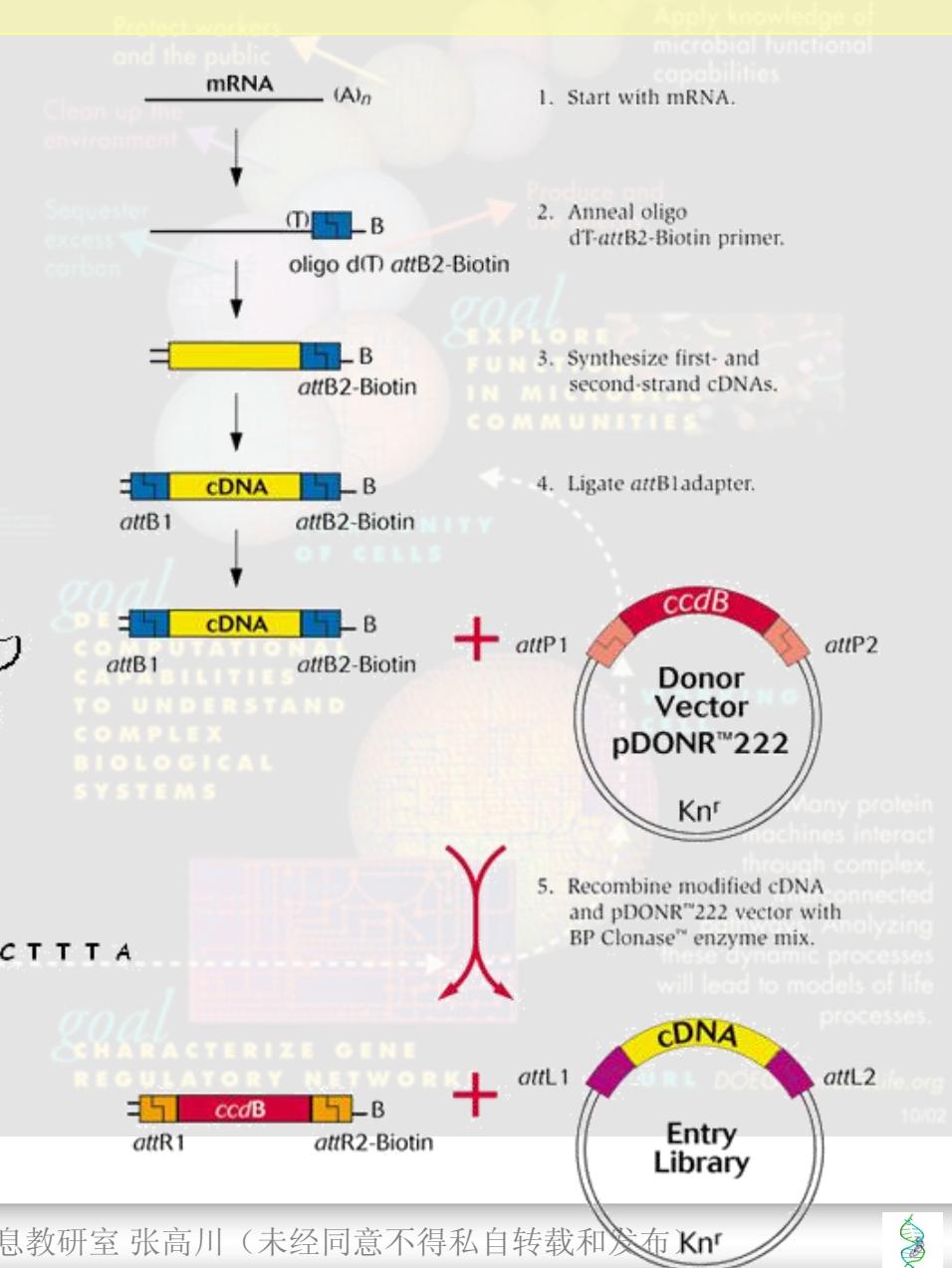
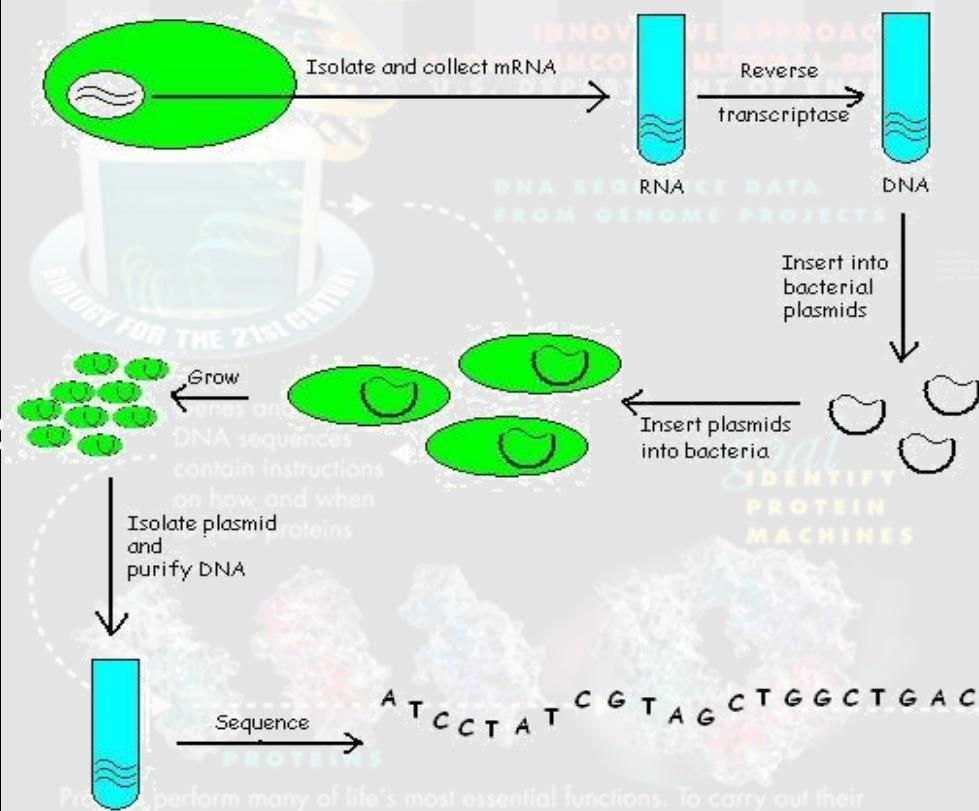
www.DOEgenomesToLife.org

10/02



cDNA Library Construction

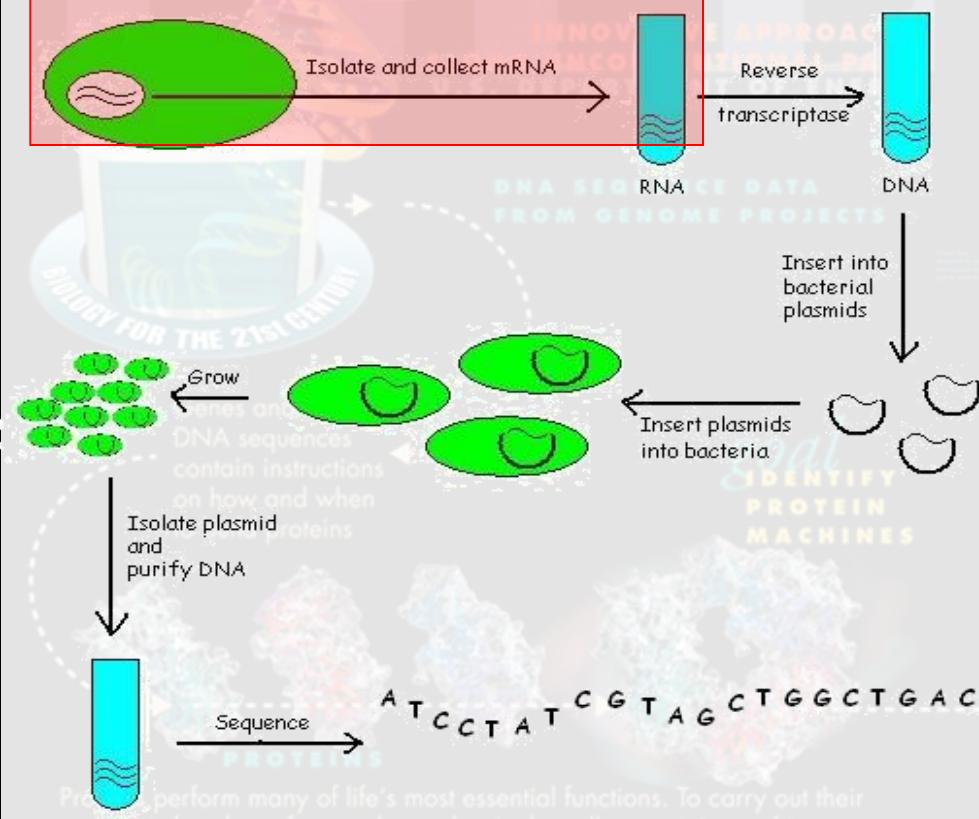
Formation of a cDNA Library



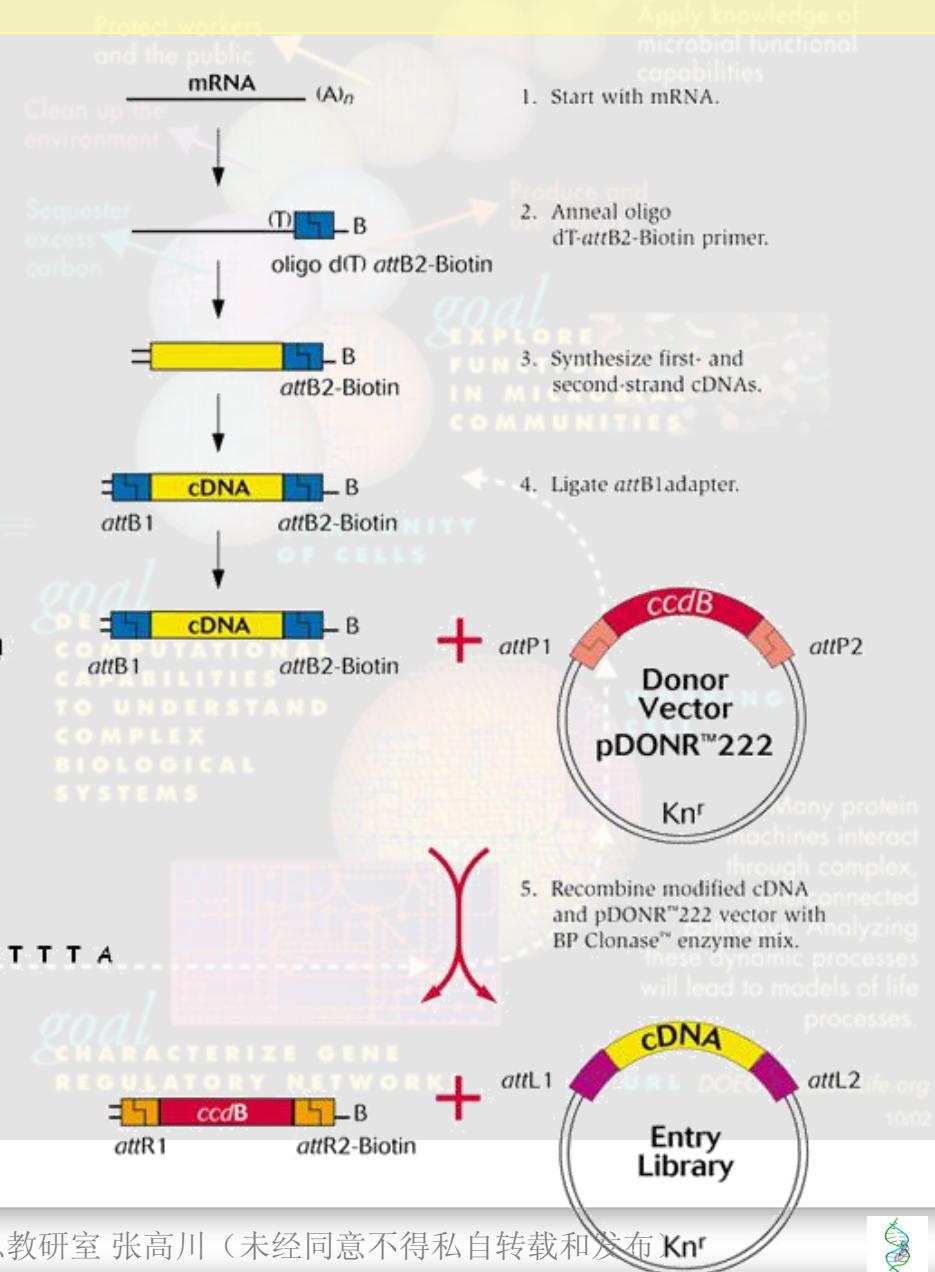
cDNA Library Construction

mRNA Extraction

Formation of a cDNA Library



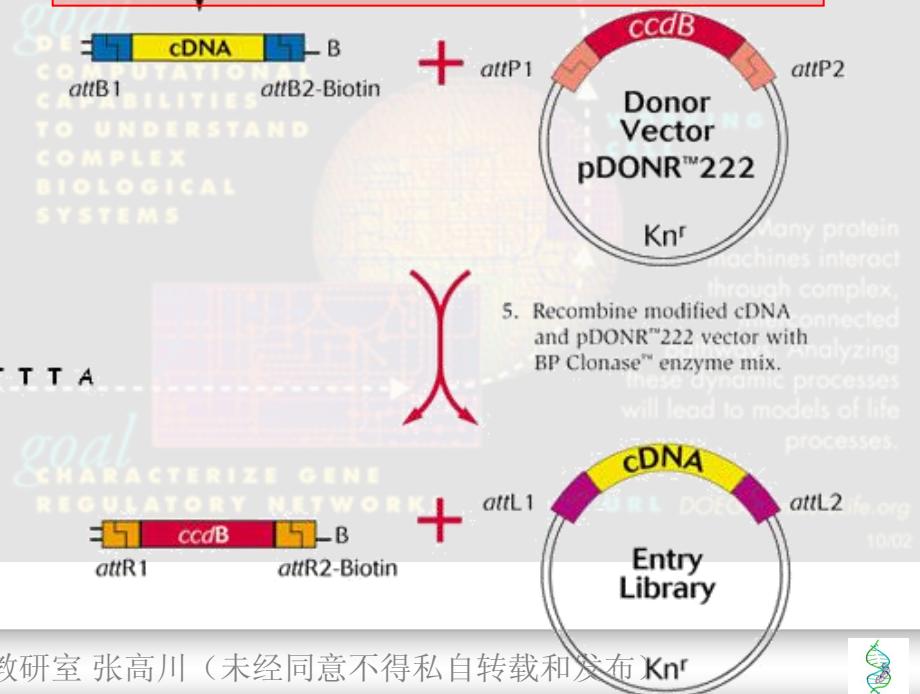
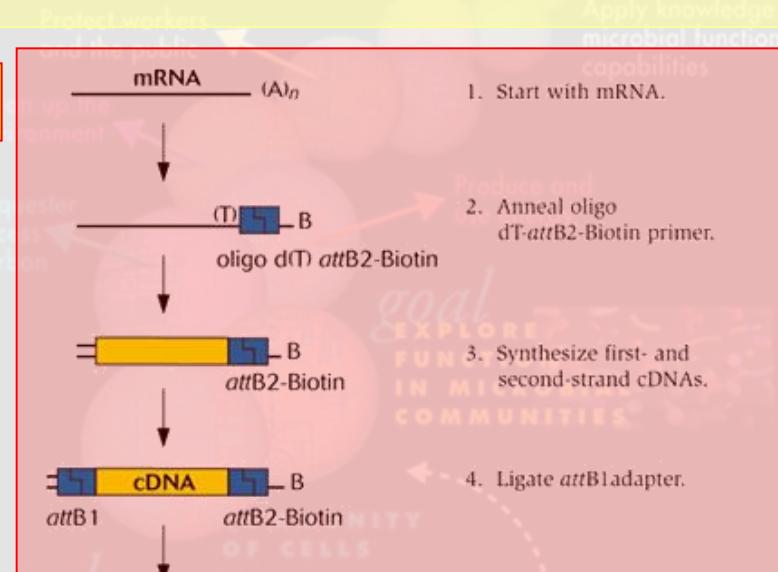
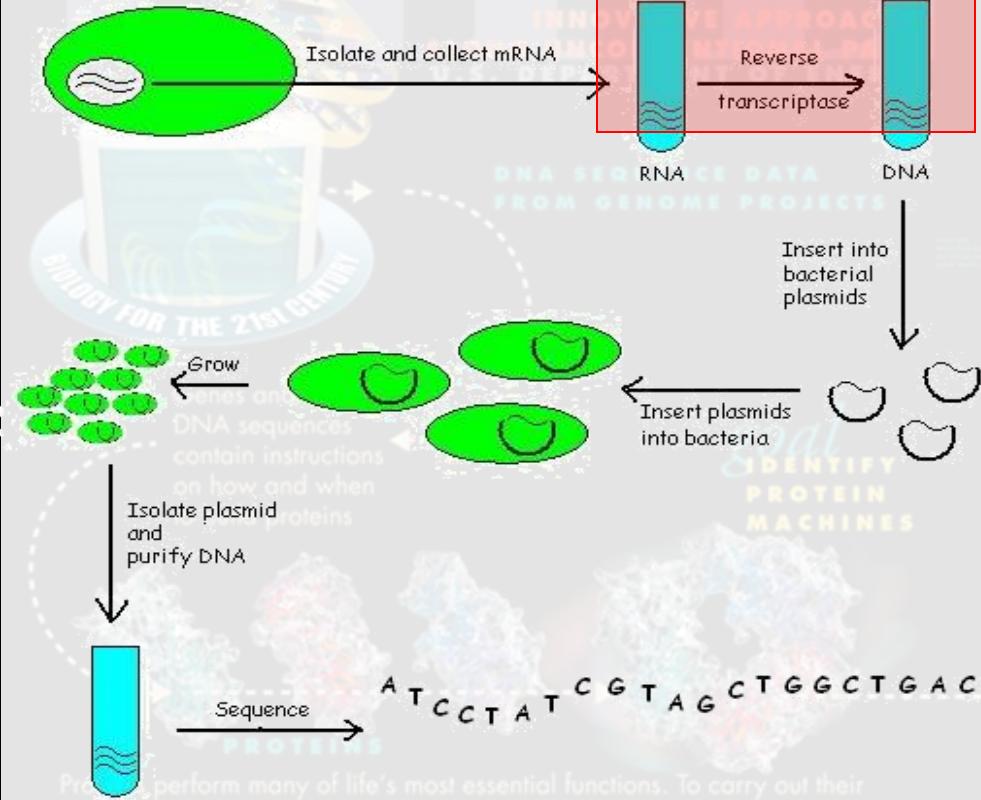
Proteins perform many of life's most essential functions. To carry out their sp. DNA roles, they often work together in the cell as protein machines.



cDNA Library Construction

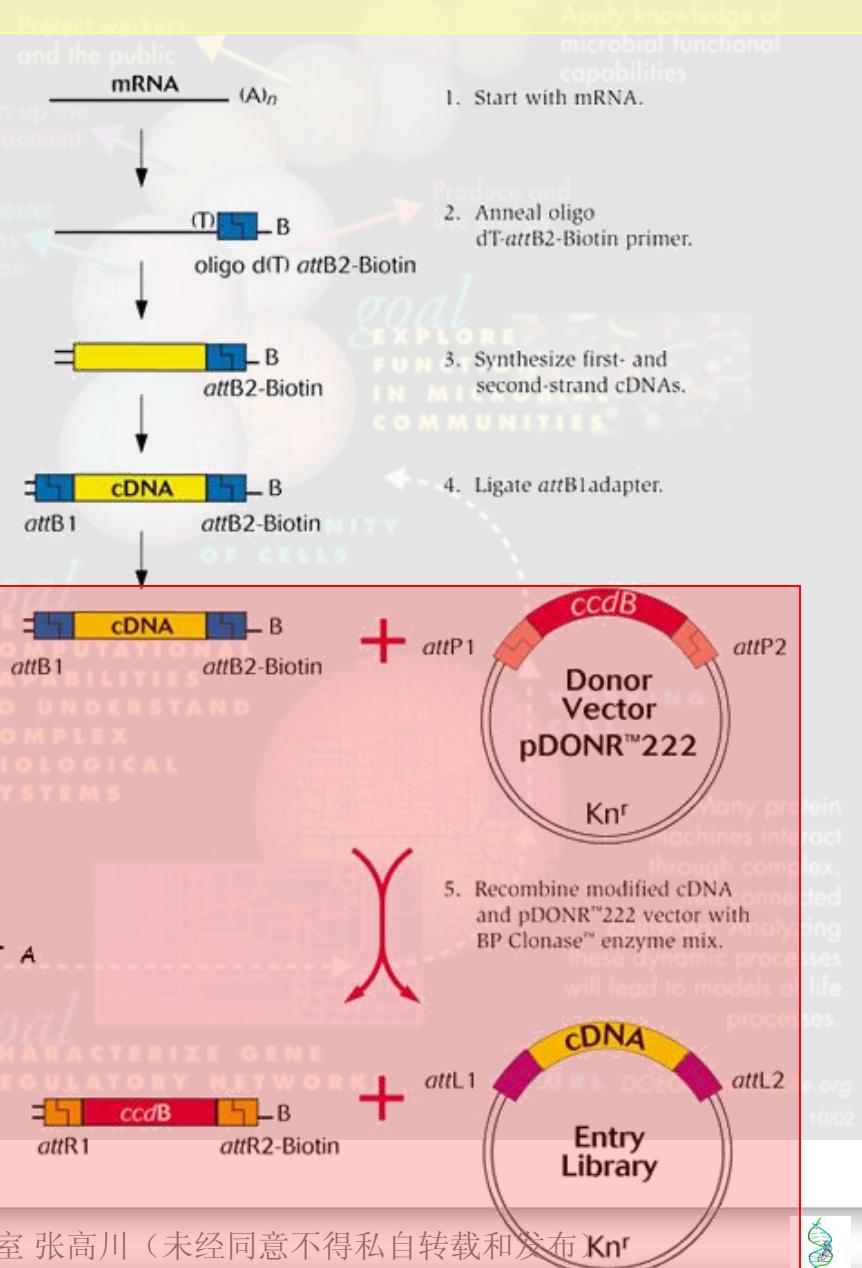
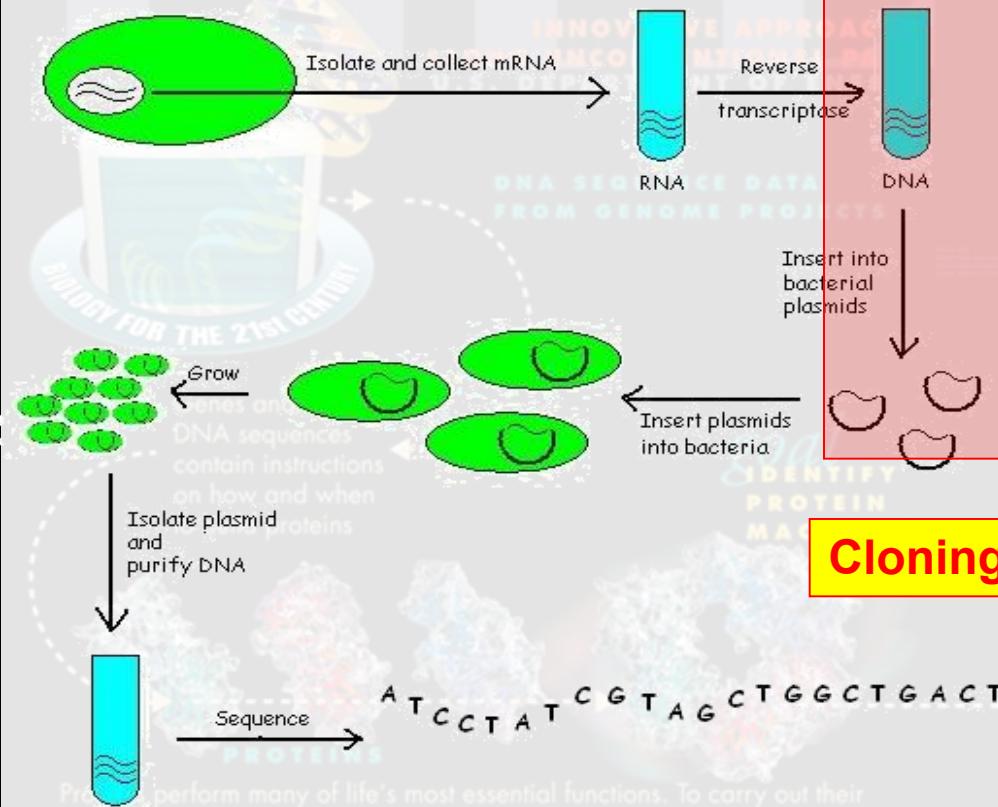
Reverse Transcription

Formation of a cDNA Library



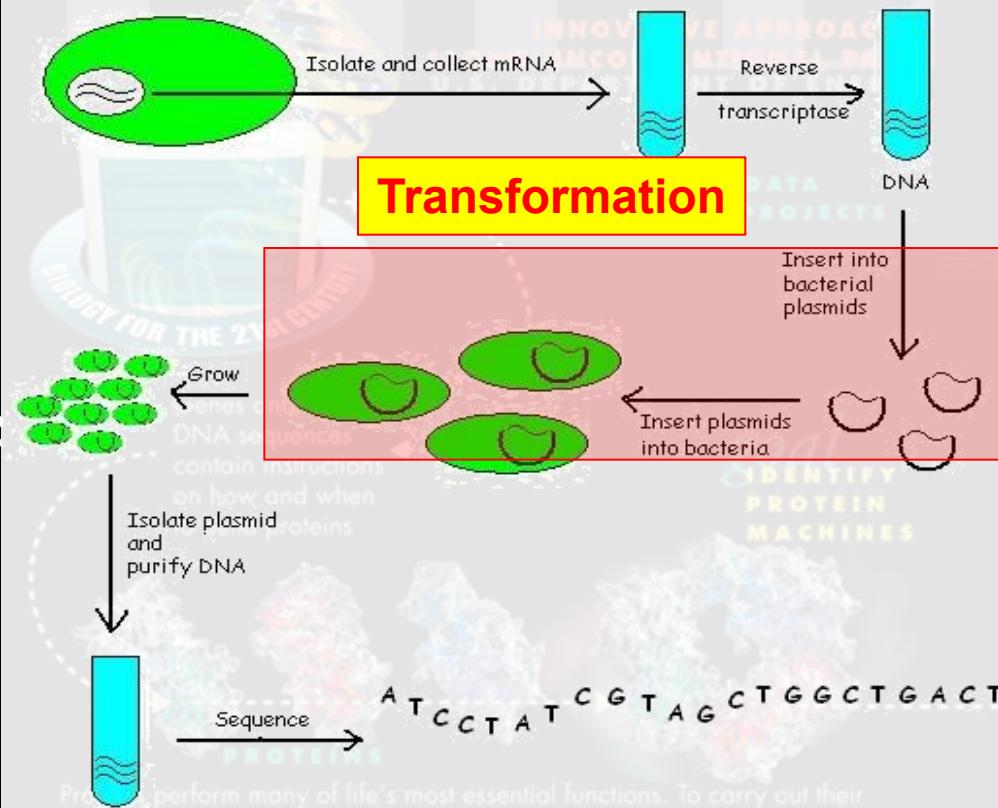
cDNA Library Construction

Formation of a cDNA Library

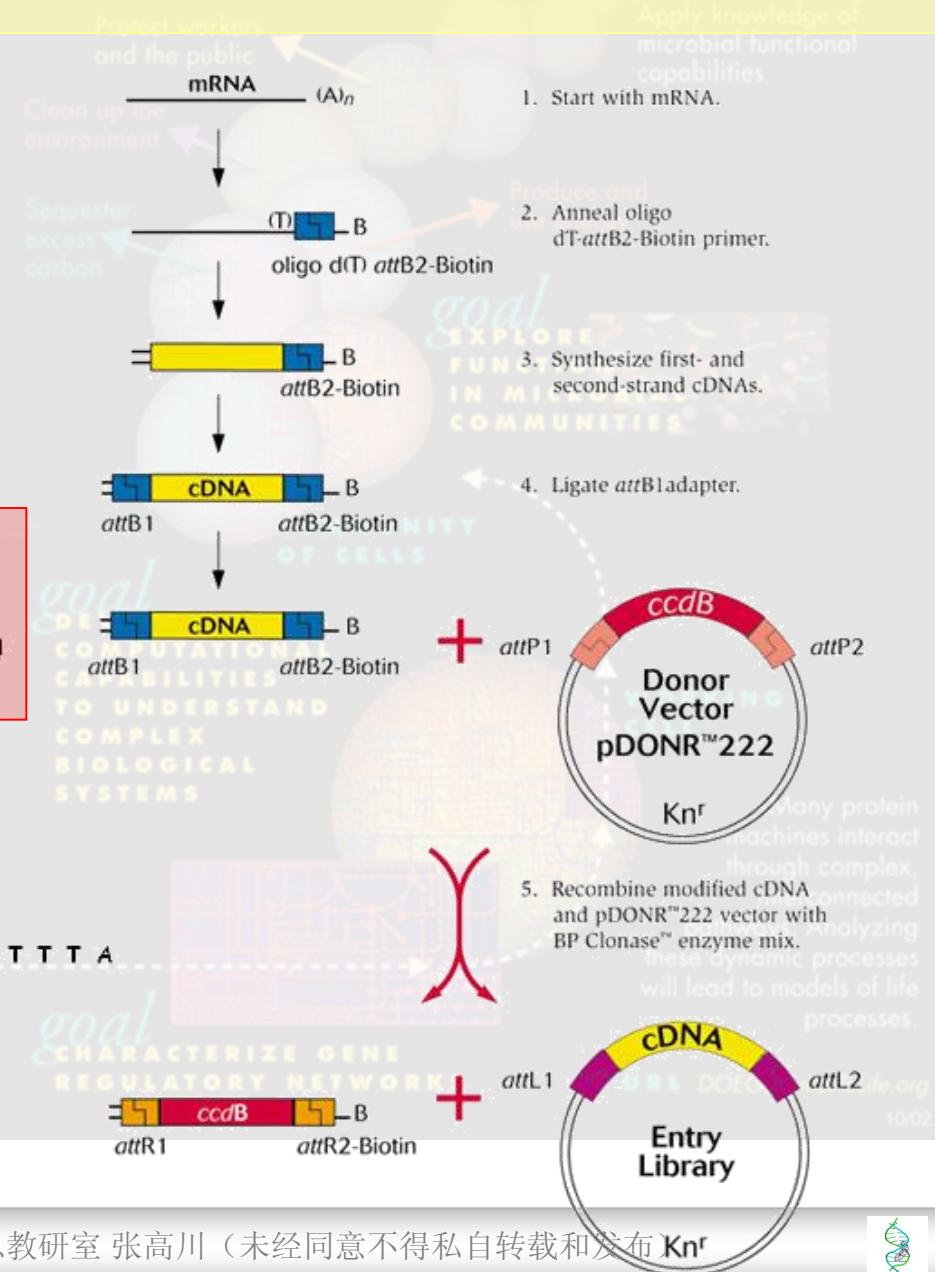


cDNA Library Construction

Formation of a cDNA Library

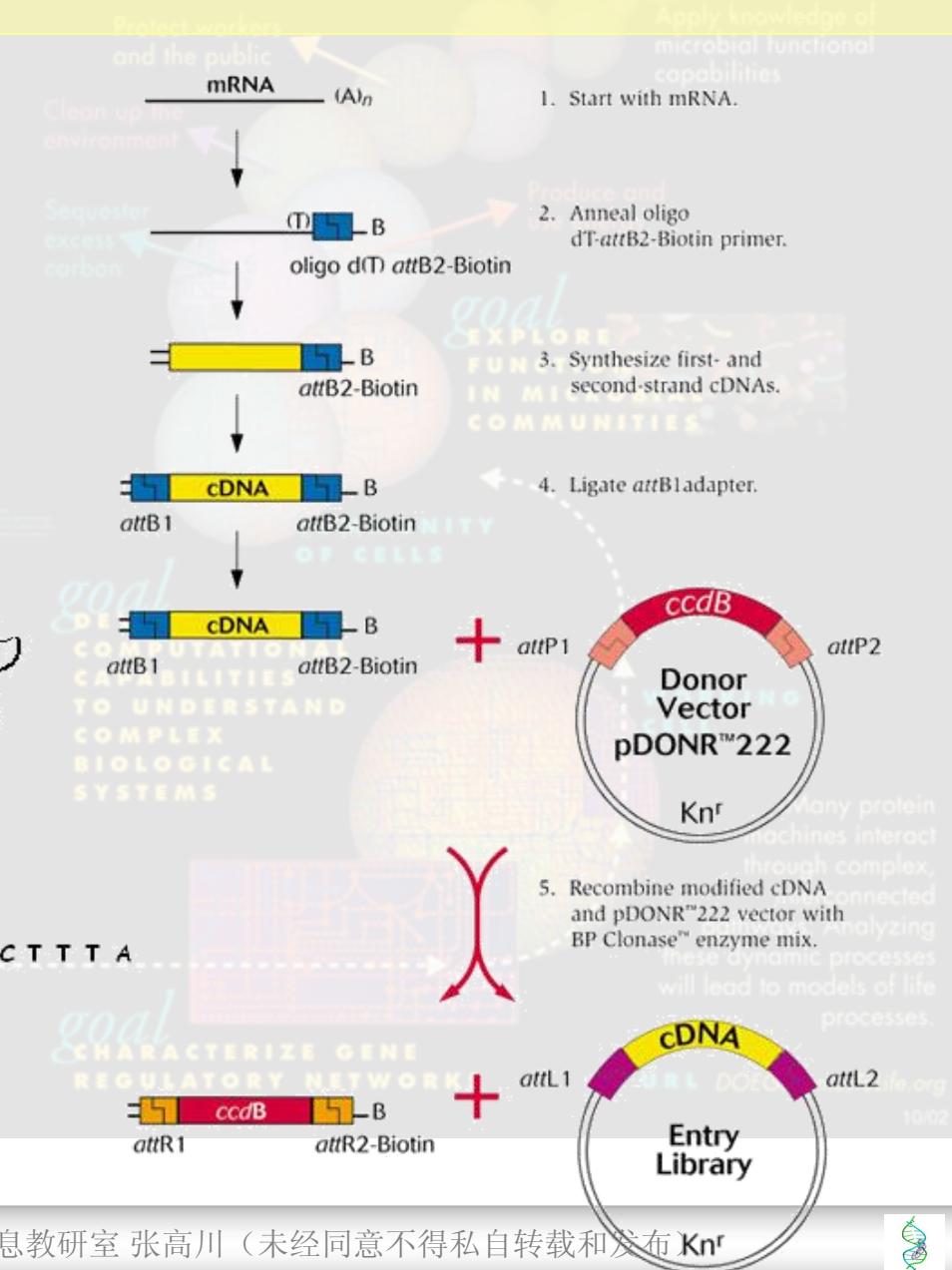
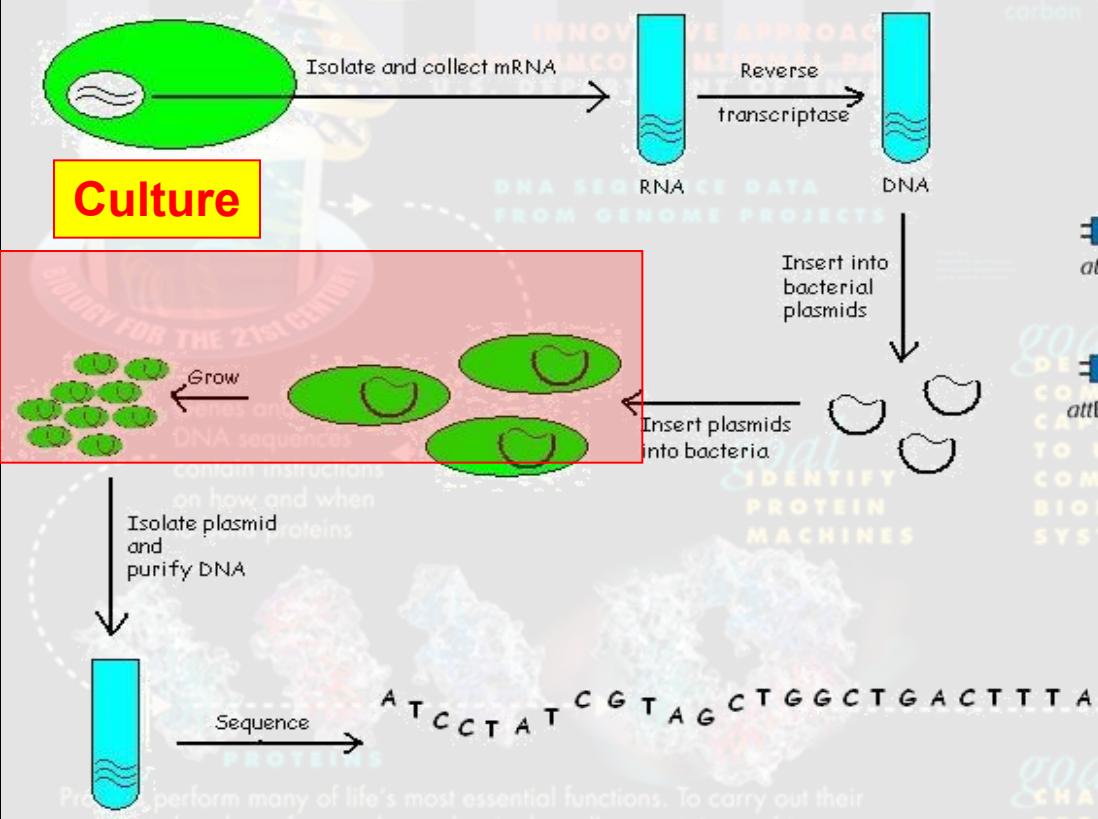


Proteins perform many of life's most essential functions. To carry out their sp. DNA roles, they often work together in the cell as protein machines.



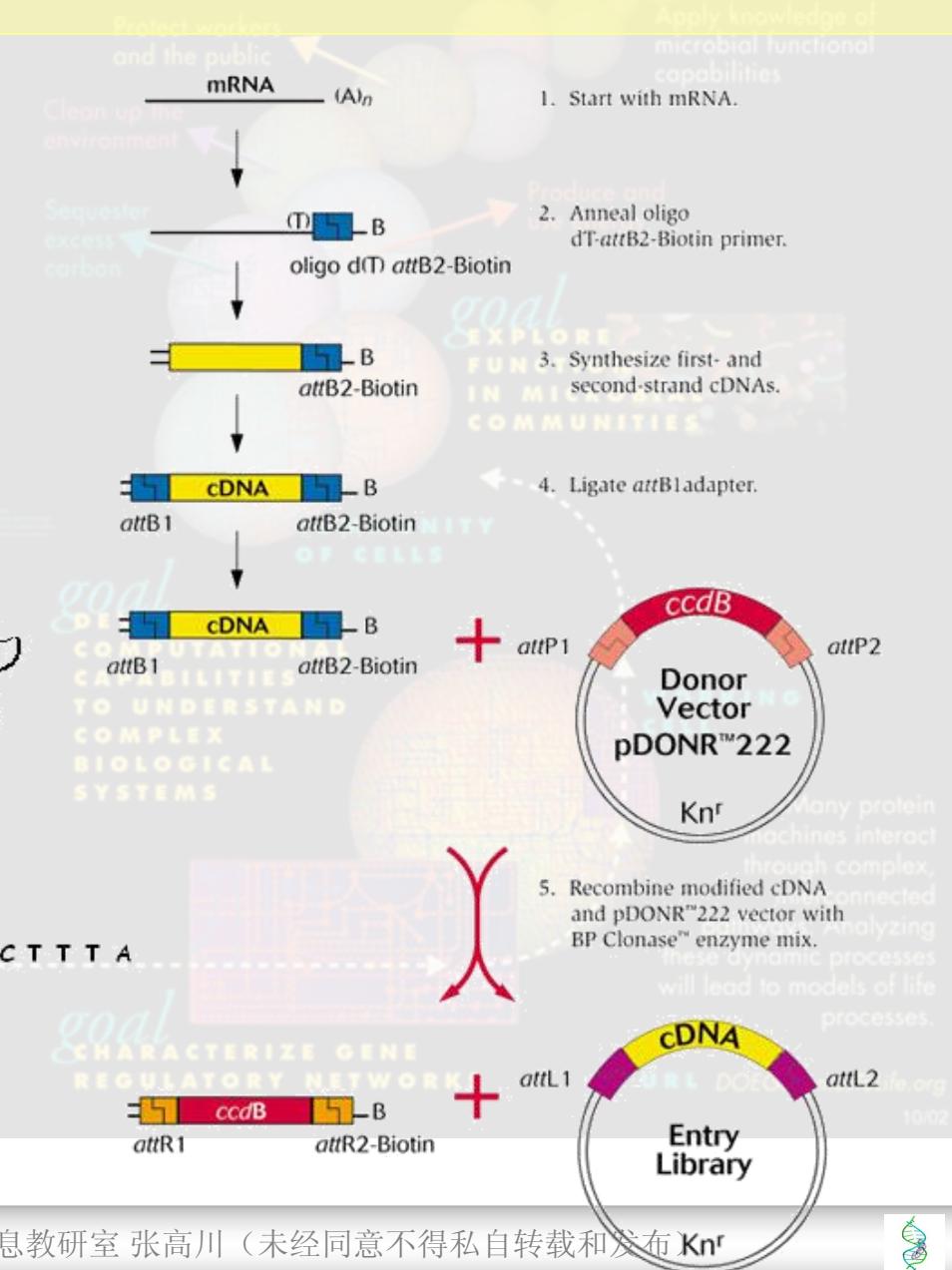
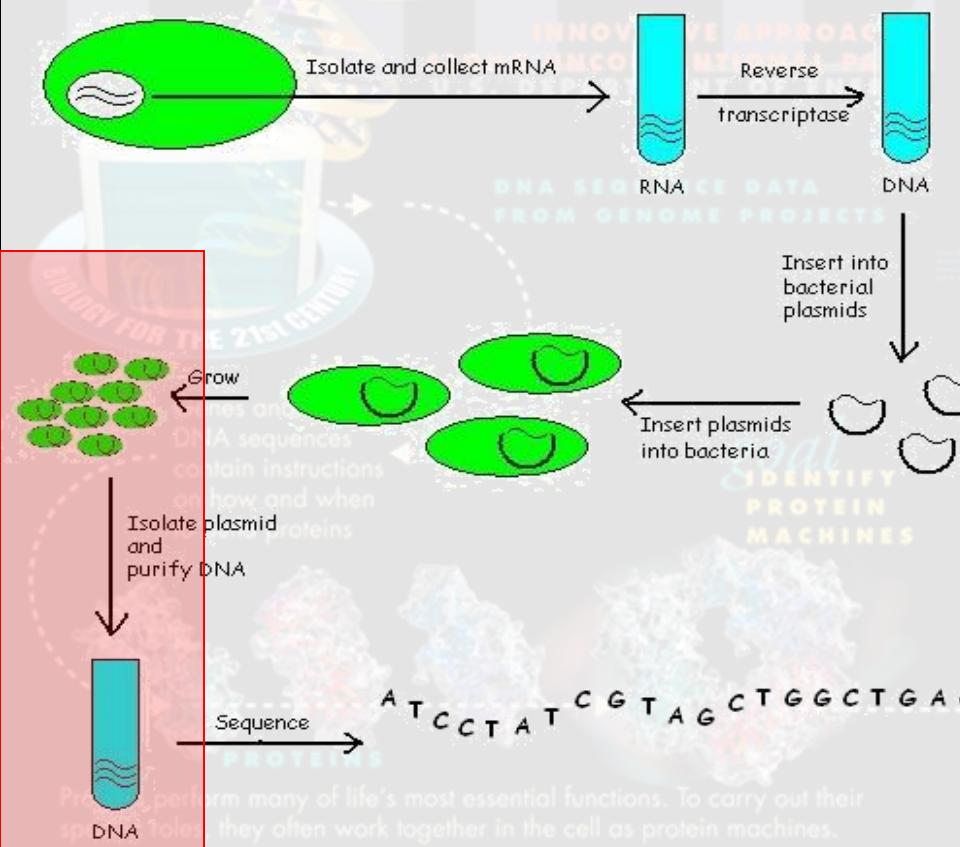
cDNA Library Construction

Formation of a cDNA Library



cDNA Library Construction

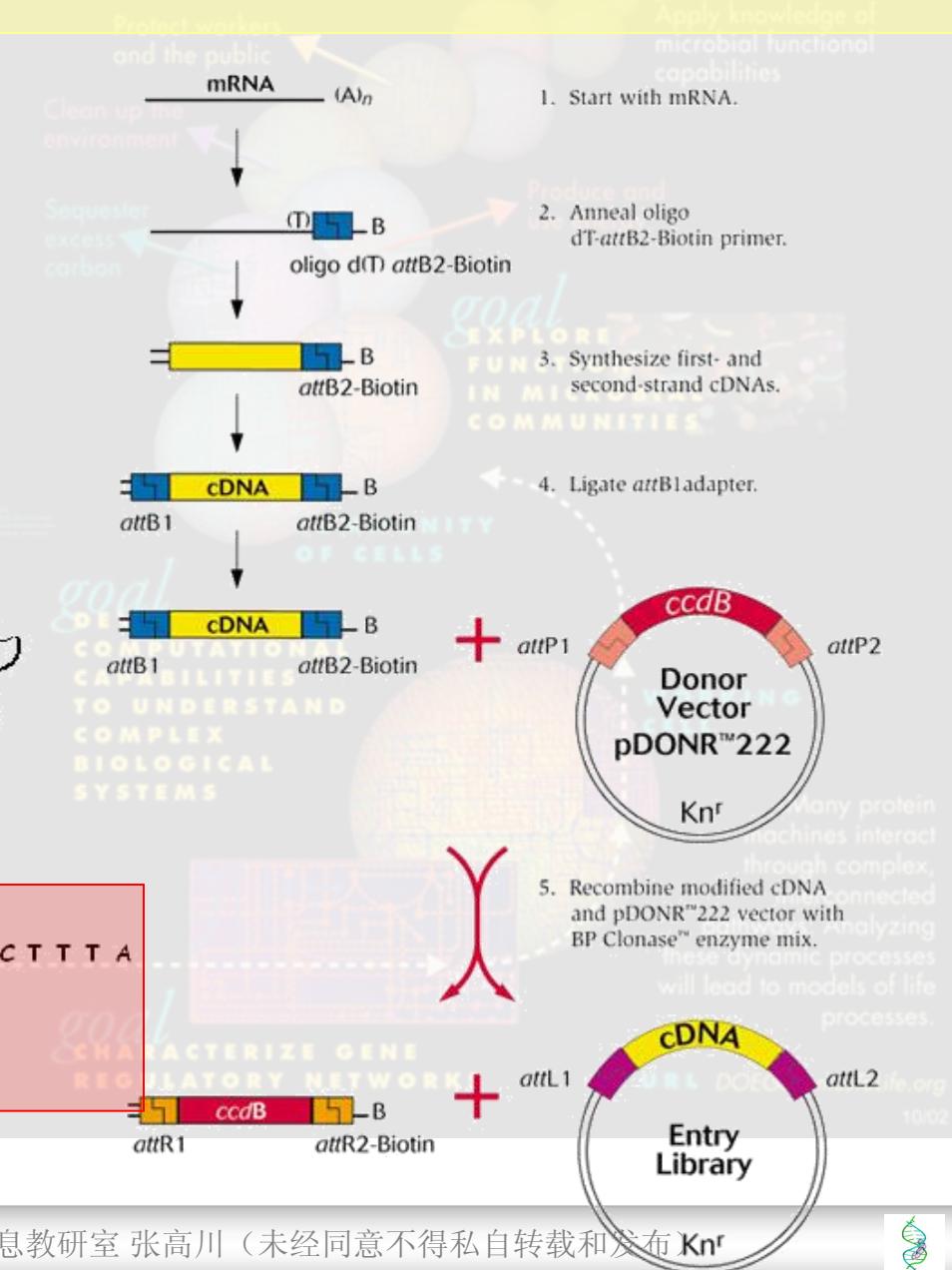
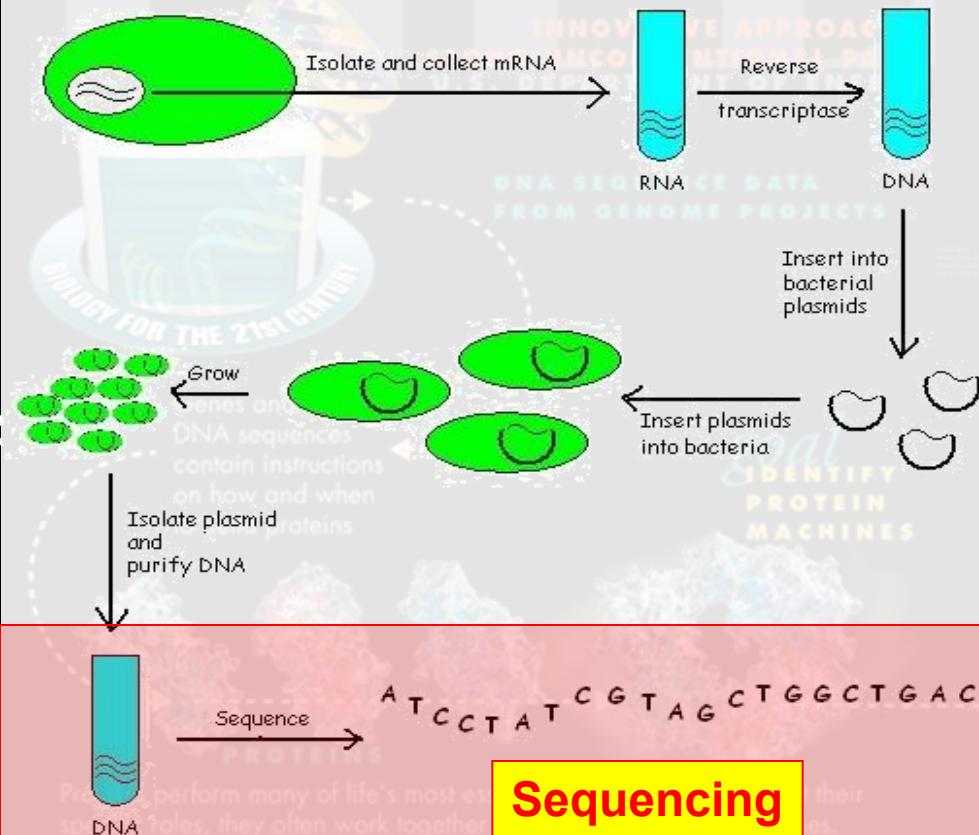
Formation of a cDNA Library



Plasmid Extraction

cDNA Library Construction

Formation of a cDNA Library



Search “cDNA library” in PubMed

NCBI Resources How To Sign in to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed cDNA library RSS Save search Advanced Help

Show additional filters Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Filters: Manage Filters

Results: 1 to 20 of 60313 << First < Prev Page 1 of 3016 Next > Last >>

[Hybrid sequencing approach applied to human fecal metagenomic clone libraries revealed clones with potential biotechnological applications.](#)
1. [Džunková M, D'Auria G, Pérez-Villarroya D, Moya A.](#)
PLoS One. 2012;7(10):e47654. doi: 10.1371/journal.pone.0047654. Epub 2012 Oct 17.
PMID: 23082187 [PubMed - in process]

[Peripheral arterial disease: a literature review.](#)
2. [Abdulhannan P, Russell DA, Homer-Vanniasinkam S.](#)
Br Med Bull. 2012 Oct 18. [Epub ahead of print]
PMID: 23080419 [PubMed - as supplied by publisher]

[Isolation and characterization of microsatellite loci from an endangered tree species, *Toona ciliata* var. *pubescens*.](#)
3. [Liu J, Sun ZX, Chen YT, Jiang JM.](#)
Genet Mol Res. 2012 Sep 19;11(AOP). [Epub ahead of print]
PMID: 23079981 [PubMed - as supplied by publisher]

[Diversion of phagosome trafficking by pathogenic *Rhodococcus equi* depends](#)

Results by year



Related searches

- cdna library construction not genomic
- cdna library screening
- brain cdna library
- cdna library review
- arabidopsis cdna library

Titles with your search terms

- Identification of microRNAs and other small regulatory RNAs us [Methods. 2008]



Search “cDNA library” in PubMed

NCBI Resources How To Sign in to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed cDNA library RSS Save search Advanced Help

Show additional filters Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Filters: Manage Filters

Results: 1 to 20 of 60313 << First < Prev Page 1 of 3016 Next > Last >>

[Hybrid sequencing approach applied to human fecal metagenomic clone libraries revealed clones with potential biotechnological applications.](#)
1. [Džunková M, D'Auria G, Pérez-Villarroya D, Moya A.](#)
PLoS One. 2012;7(10):e47654. doi: 10.1371/journal.pone.0047654. Epub 2012 Oct 17.
PMID: 23082187 [PubMed - in process]

[Peripheral arterial disease: a literature review.](#)
2. [Abdulhannan P, Russell DA, Homer-Vanniasinkam A.](#)
Br Med Bull. 2012 Oct 18. [Epub ahead of print]
PMID: 23080419 [PubMed - as supplied by publisher]

[Isolation and characterization of microsatellite loci from an endangered tree species, *Toona ciliata* var. *pubescens*.](#)
3. [Liu J, Sun ZX, Chen YT, Jiang JM.](#)
Genet Mol Res. 2012 Sep 19;11(AOP). [Epub ahead of print]
PMID: 23079981 [PubMed - as supplied by publisher]

[Diversion of phagosome trafficking by pathogenic *Rhodococcus equi* depends](#)

Results by year

Related searches

cdna library construction not genomic
cdna library screening
brain cdna library
cdna library review
arabidopsis cdna library

Titles with your search terms

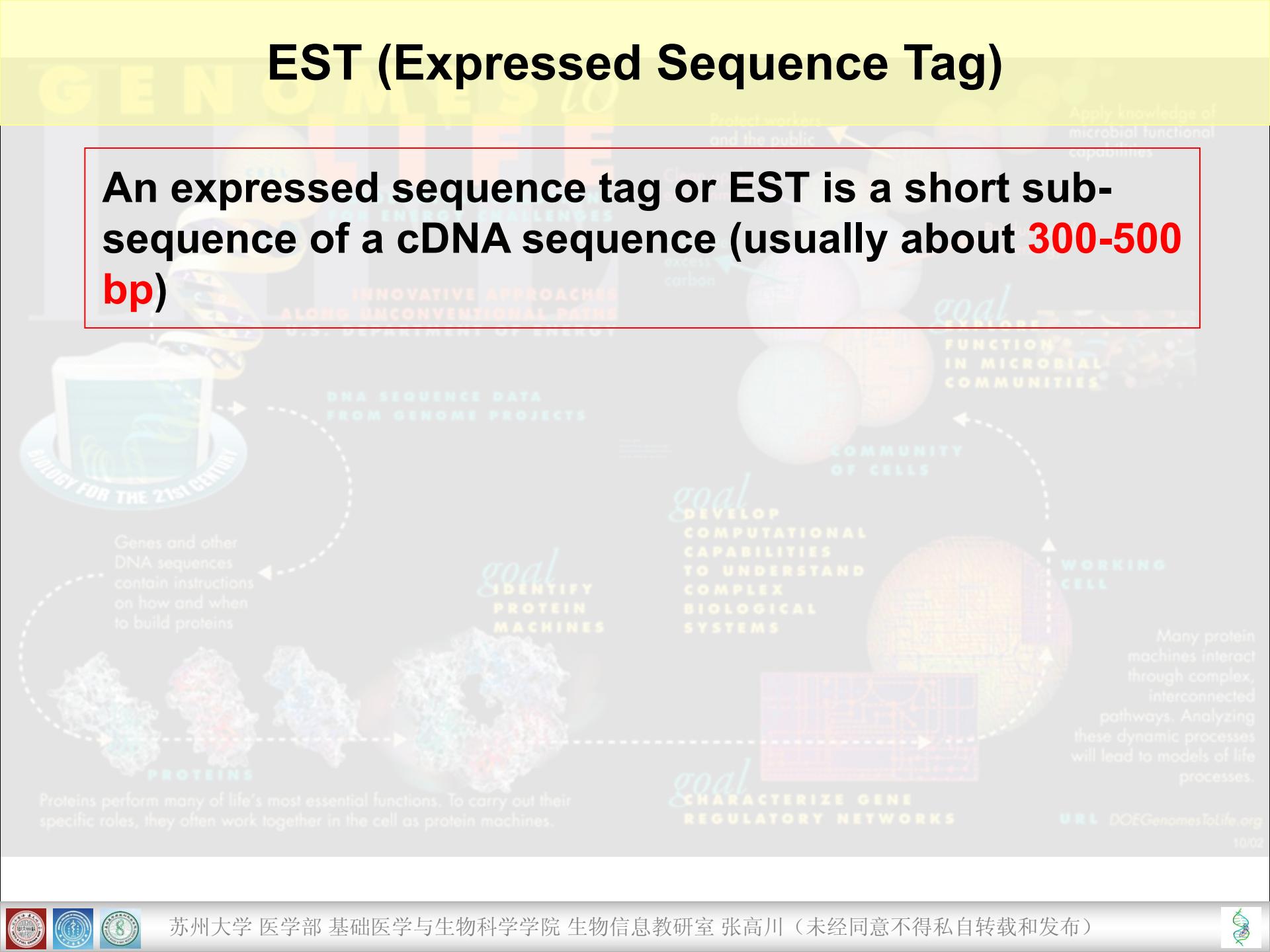
Identification of microRNAs and other small regulatory RNAs us [Methods. 2008]

1980s, 开始兴起



EST (Expressed Sequence Tag)

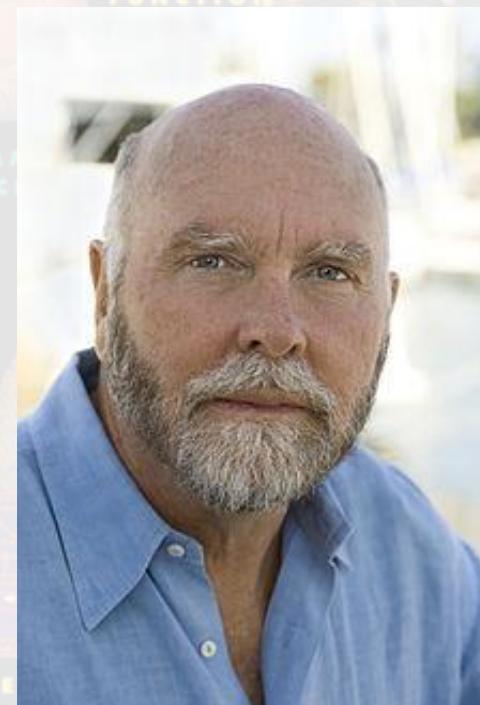
An expressed sequence tag or EST is a short subsequence of a cDNA sequence (usually about **300-500 bp**)



EST (Expressed Sequence Tag)

An expressed sequence tag or EST is a short subsequence of a cDNA sequence (usually about 300-500 bp)

1. 90年代初, **Graig Venter** 提出了EST的概念, 并测定了609条人脑组织的EST, 宣布了cDNA大规模测序的时代的开始 (**Adams et al., 1991, Science**)。
1. 1993年前ESTs数据收录于GenBank, EBI和DDBJ。
1. 1993年NCBI(National Center of Biotechnology Information)建立了一个专门的EST数据库dbEST来保存和收集所有的EST数据。

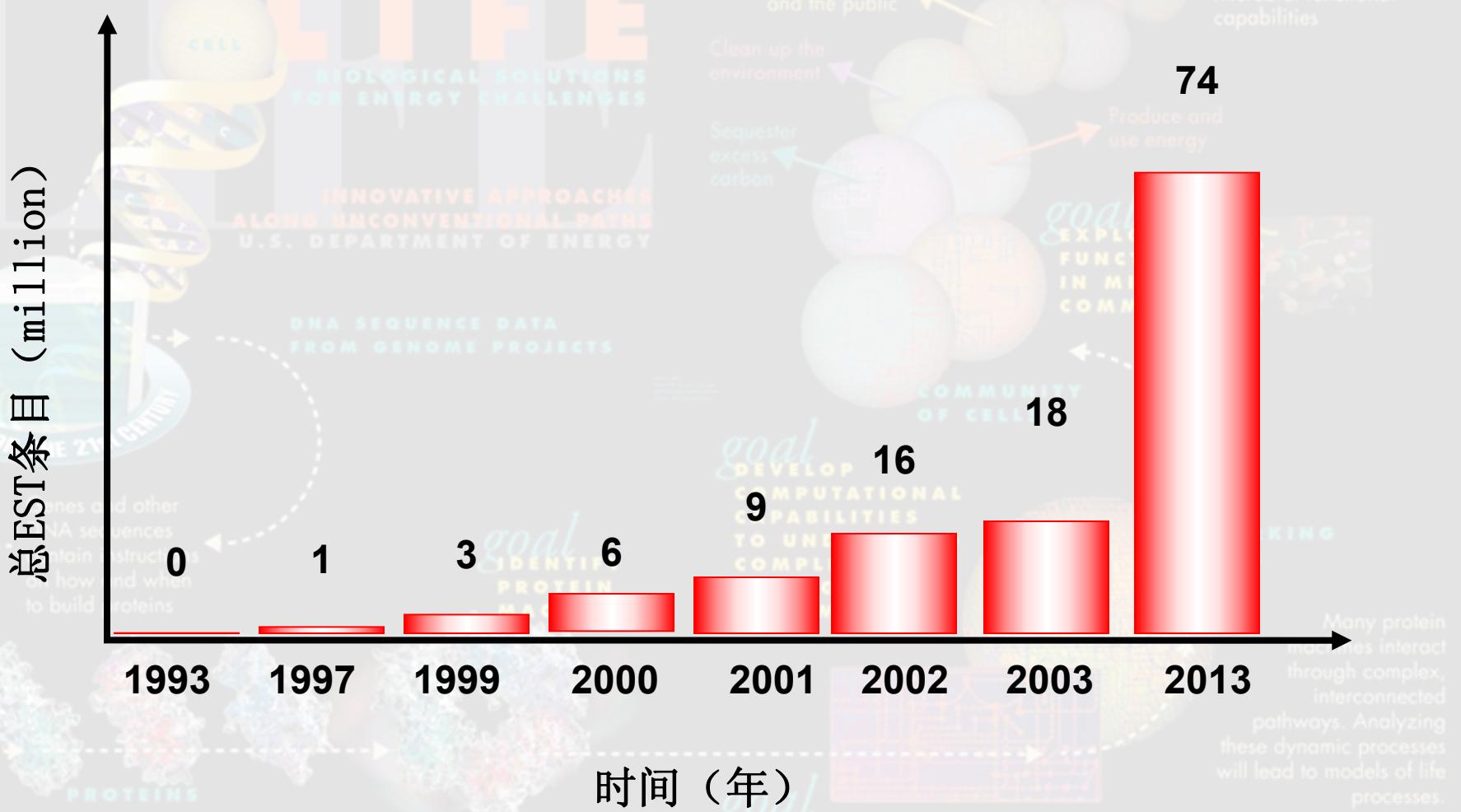


<http://www.jcvi.org/>

http://en.wikipedia.org/wiki/Craig_Venter



dbEST中数据量的增长



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

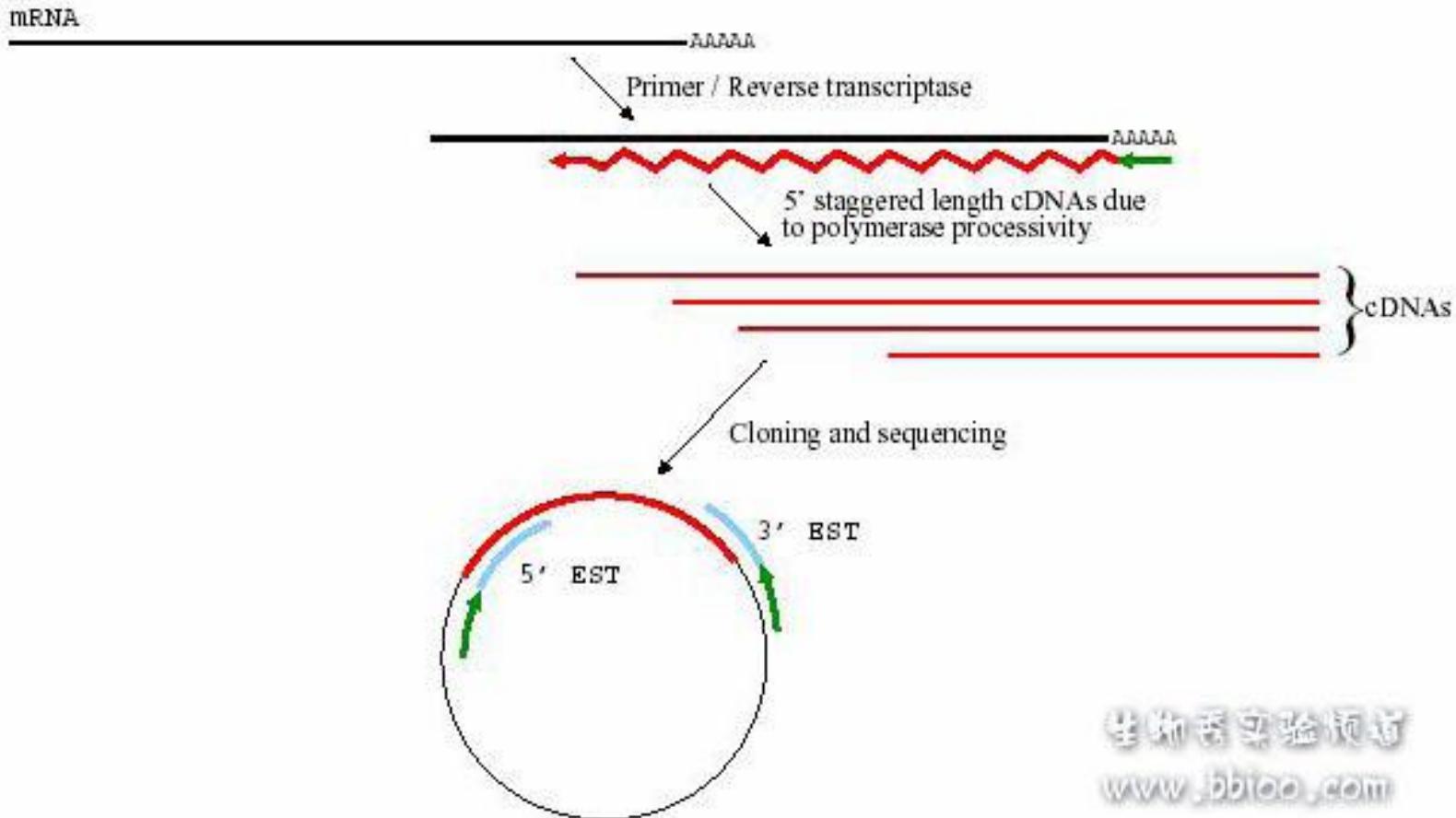
URL: DOEGenomesToLife.org

10/02



EST sequencing

ESTs (Expressed Sequence tags) 是从已建好的cDNA库中随机取出一个克隆，从5'末端或3'末端对插入的cDNA片段进行一轮单向自动测序，所获得的约500bp的一段cDNA序列

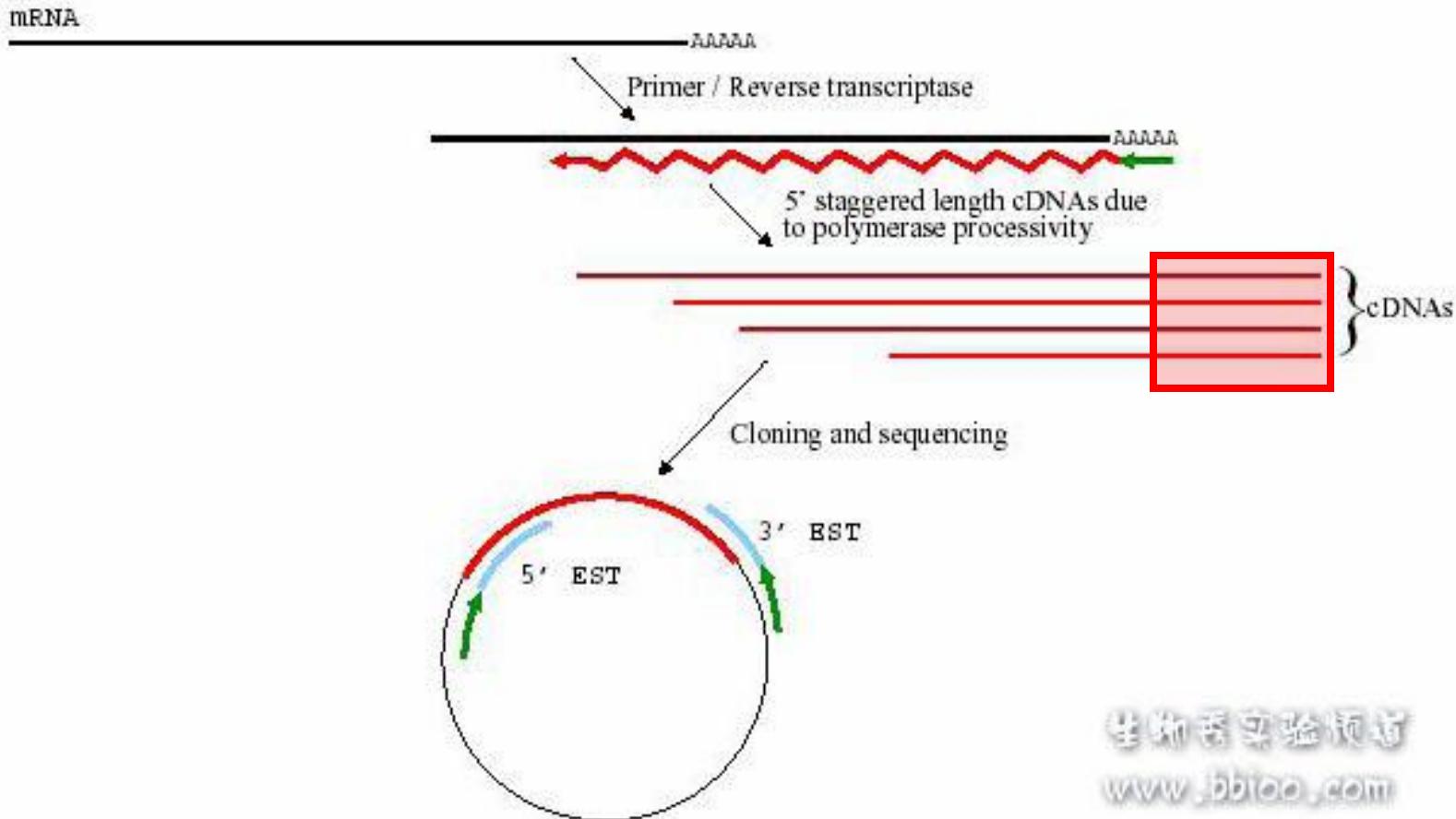


生物信息学
www.bb100.com



EST sequencing

ESTs (Expressed Sequence tags) 是从已建好的cDNA库中随机取出一个克隆，从5'末端或3'末端对插入的cDNA片段进行一轮单向自动测序，所获得的约500bp的一段cDNA序列

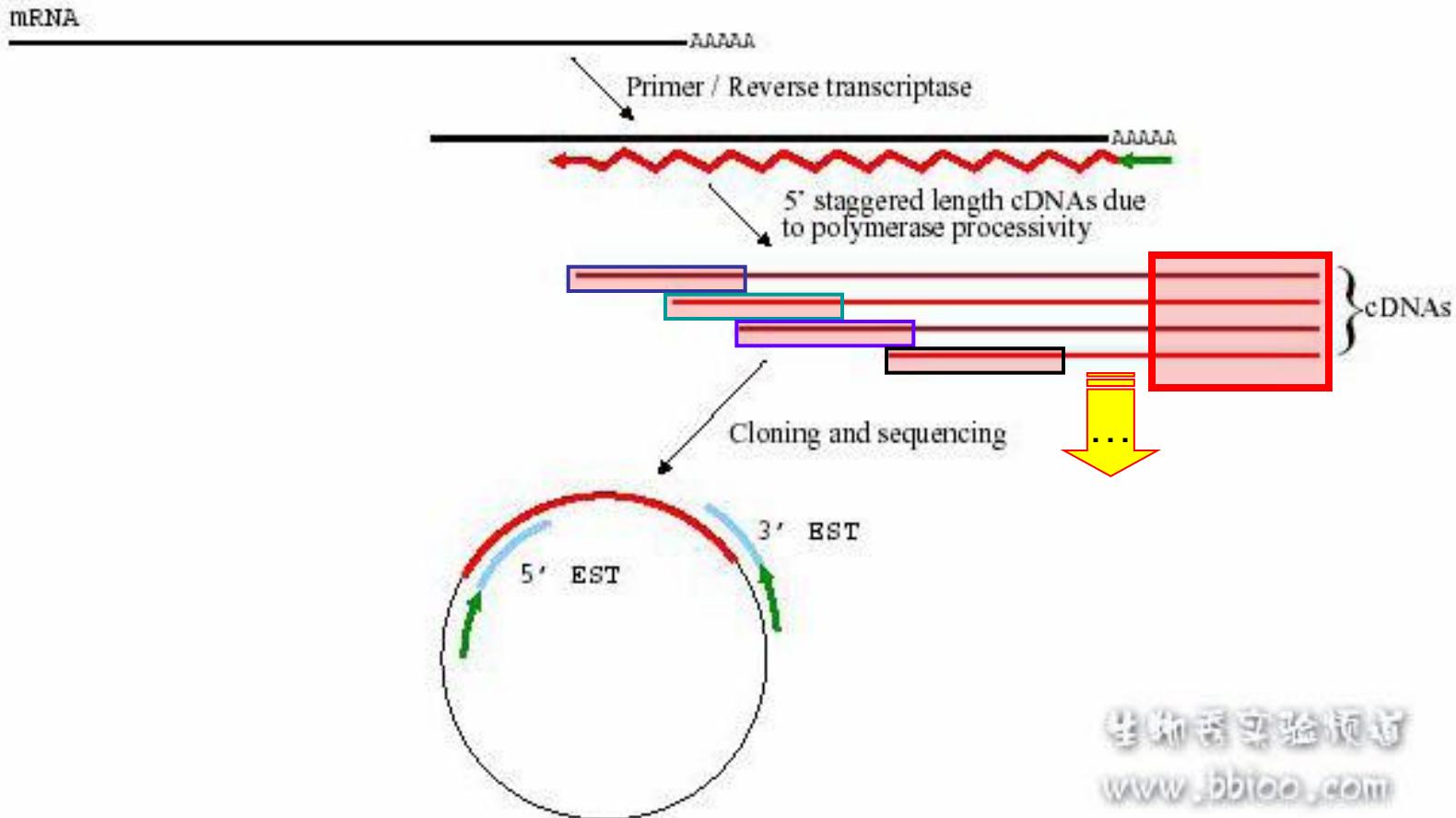


生物信息学
www.bb100.com



EST sequencing

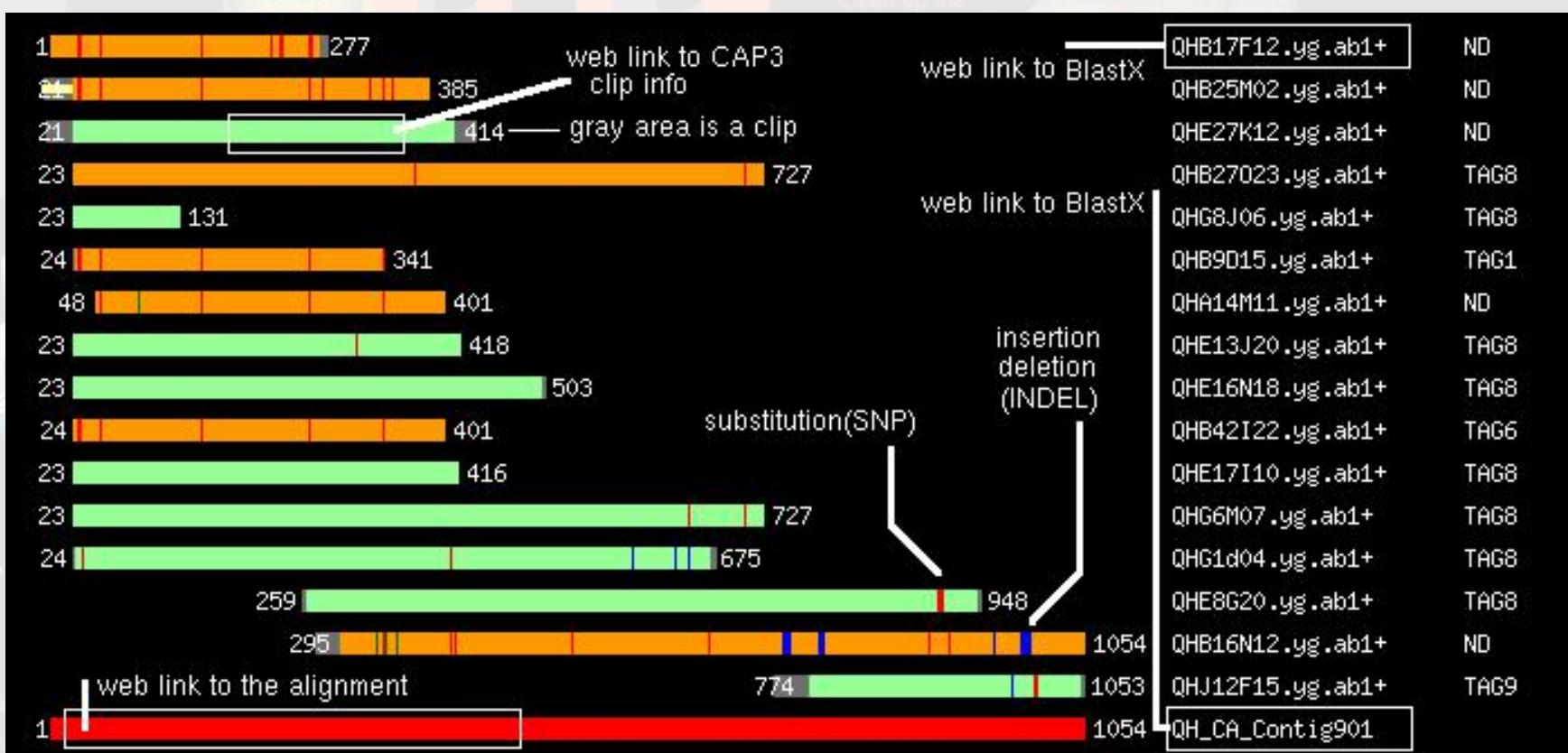
ESTs (Expressed Sequence tags) 是从已建好的cDNA库中随机取出一个克隆，从5'末端或3'末端对插入的cDNA片段进行一轮单向自动测序，所获得的约500bp的一段cDNA序列



生物信息学
www.bb100.com



EST assembly and contigs



PROTEINS

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

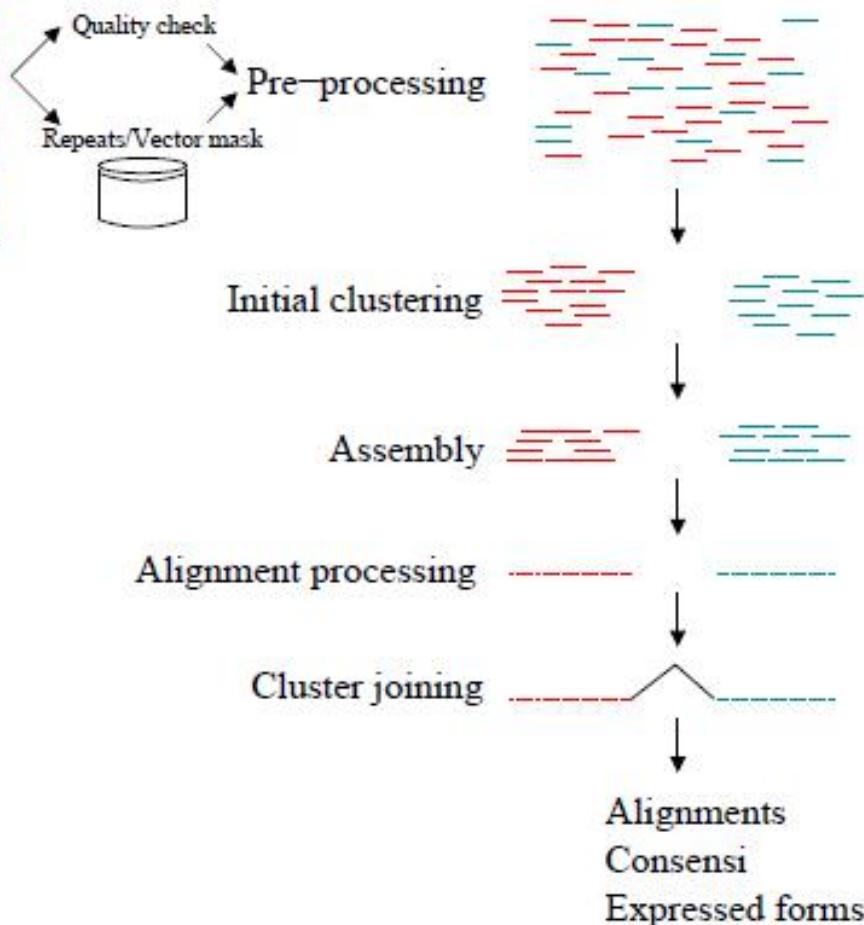
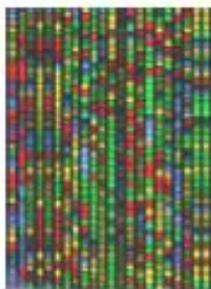


EST测序和组装

Project workers
2000-2001

Apply knowledge of
microbial functional

25



Unigene

**TIGR
STACK**

Many protein machines interact through complex, interconnected ways. Analyzing dynamic processes leads to models of life processes.

GenomesToLife.org

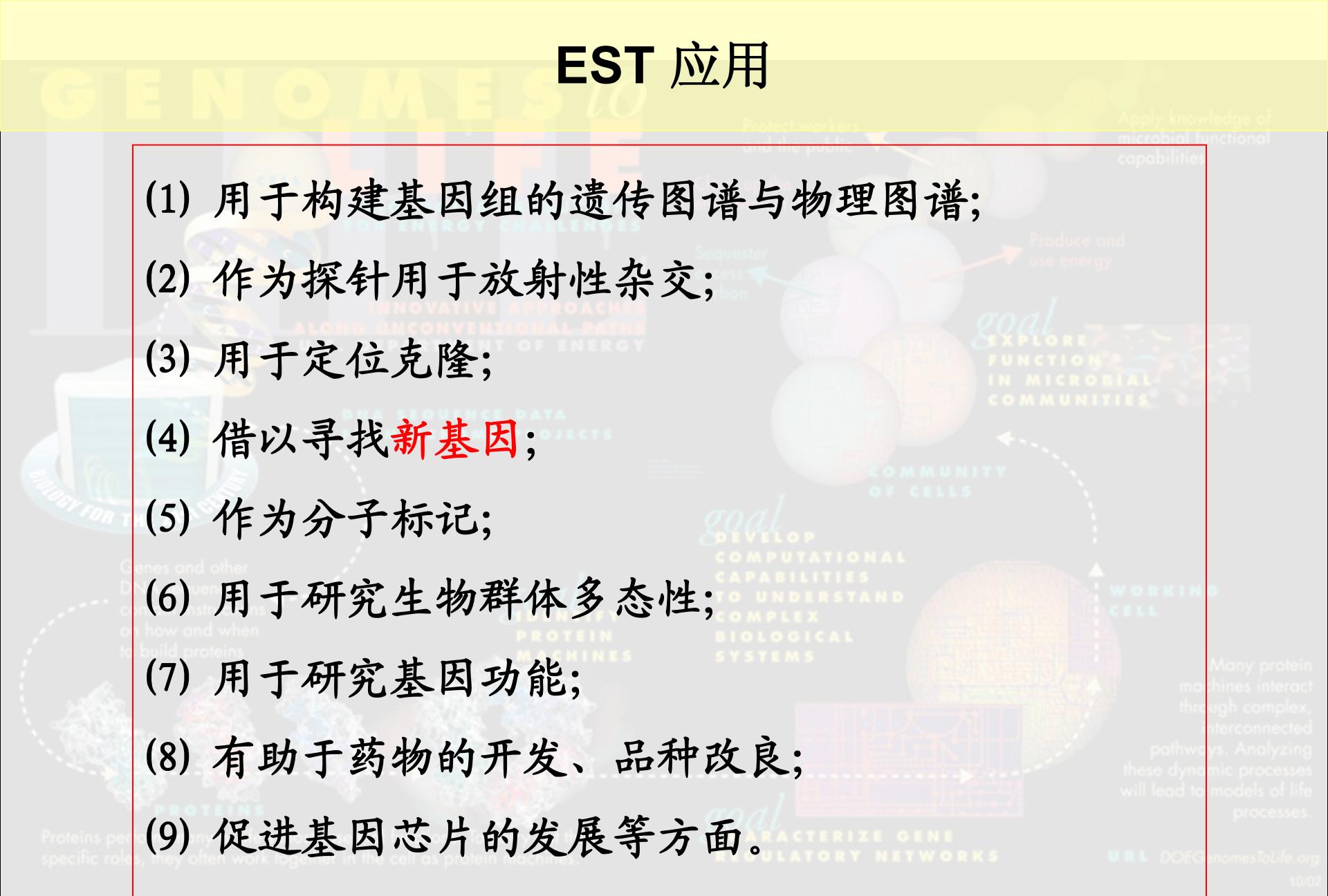
10/02

Proteins perform specific roles



EST 应用

- (1) 用于构建基因组的遗传图谱与物理图谱;
- (2) 作为探针用于放射性杂交;
- (3) 用于定位克隆;
- (4) 借以寻找新基因;
- (5) 作为分子标记;
- (6) 用于研究生物群体多态性;
- (7) 用于研究基因功能;
- (8) 有助于药物的开发、品种改良;
- (9) 促进基因芯片的发展等方面。



EST与基因识别

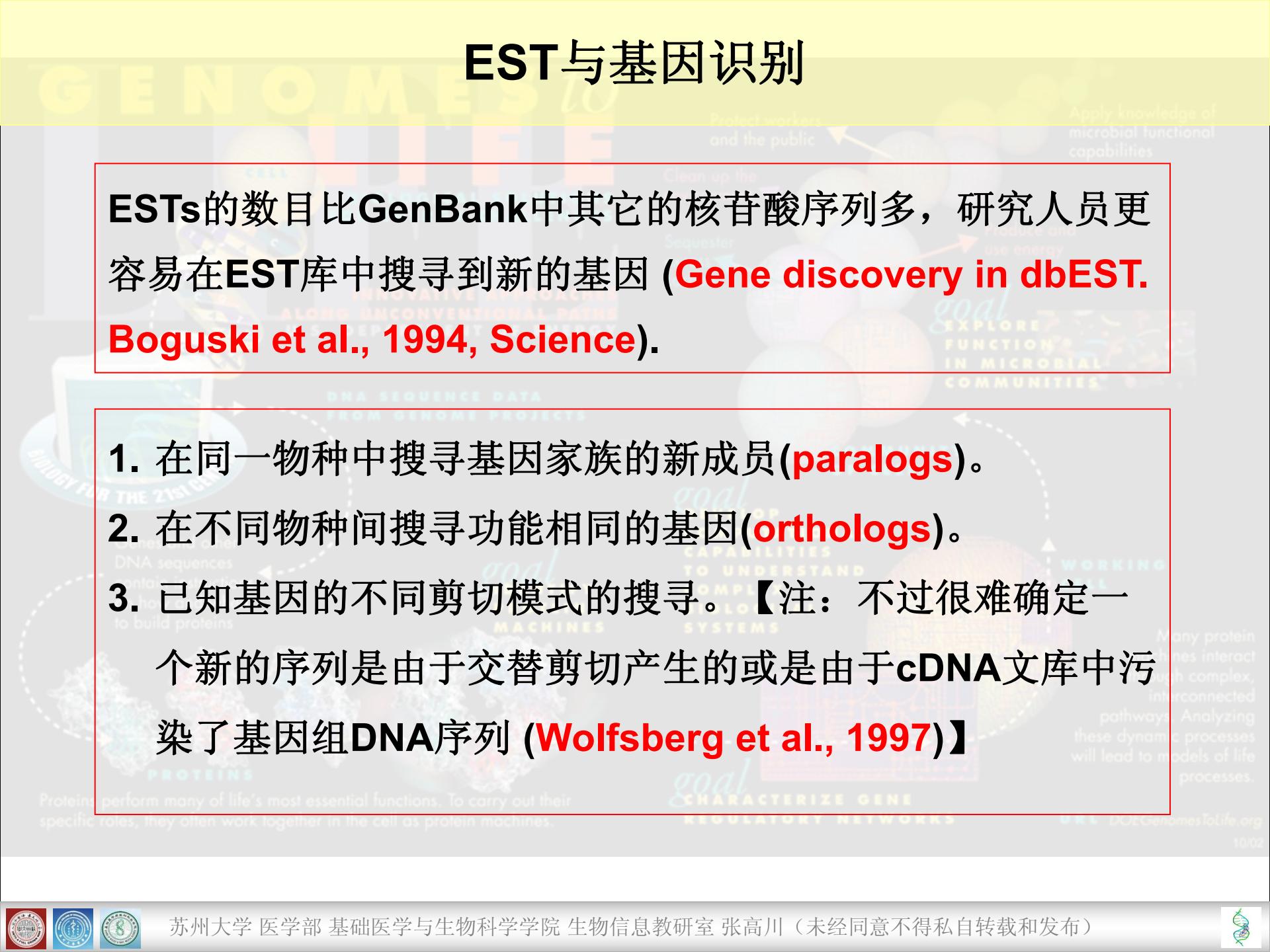
ESTs的数目比GenBank中其它的核苷酸序列多，研究人员更容易在EST库中搜寻到新的基因 (**Gene discovery in dbEST. Boguski et al., 1994, Science**).



EST与基因识别

ESTs的数目比GenBank中其它的核苷酸序列多，研究人员更容易在EST库中搜寻到新的基因 (**Gene discovery in dbEST. Boguski et al., 1994, Science**).

1. 在同一物种中搜寻基因家族的新成员(**paralogs**)。
2. 在不同物种间搜寻功能相同的基因(**orthologs**)。
3. 已知基因的不同剪切模式的搜寻。【注：不过很难确定一个新的序列是由于交替剪切产生的或是由于cDNA文库中污染了基因组DNA序列 (**Wolfsberg et al., 1997**)】



Search “expressed sequence tag[tiab]” in PubMed

Project works... Apply knowledge of microbiology and

NCBI Resources How To Sign in to NCBI

PubMed expressed sequence tag[tiab] Search Help

RSS Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Filters: Manage Filters

Text availability Abstract available Free full text available Full text available

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

Article types Clinical Trial Randomized Controlled Trial Review Systematic Reviews more ...

Results: 1 to 20 of 2814 << First < Prev Page 1 of 141 Next > Last >>

[Benchmark comparison of ab initio microRNA identification methods and software.](#)
1. Hu LL, Huang Y, Wang QC, Zou Q, Jiang Y.
Genet Mol Res. 2012 Oct 17;11(AOP). [Epub ahead of print]
PMID: 23096922 [PubMed - as supplied by publisher]
[Related citations](#)

[Immune gene discovery by expressed sequence tag \(EST\) analysis of haemocytes in the ridgetail white prawn Exopalaemon carinicauda.](#)
2. Duan Y, Liu P, Li J, Li J, Chen P.
Fish Shellfish Immunol. 2012 Oct 20. pii: S1050-4648(12)00391-9. doi:
10.1016/j.fsi.2012.10.026. [Epub ahead of print]
PMID: 23092732 [PubMed - as supplied by publisher]
[Related citations](#)

[Scratching the itch: new tools to advance understanding of scabies.](#)
3. Mounsey KE, McCarthy JS, Walton SF.
Trends Parasitol. 2012 Oct 19. pii: S1471-4922(12)00166-3. doi: 10.1016/j.pt.2012.09.006.
[Epub ahead of print]
PMID: 23088958 [PubMed - as supplied by publisher]
[Related citations](#)

Results by year

Titles with your search terms

A hitchhiker's guide to **expressed sequence tag** [Brief Bioinform. 2007]

Generation of an **expressed sequence t**: [Plant Biol (Stuttg). 2012]

Identification and validation of Asteraceae miRNAs by tr [Gene. 2012]

See more...

769 free full-text articles in PubMed Central

SNP Discovery with EST and NextGen Sequencing in Swic [PLoS One. 2012]

SUZUKE MEDICAL COLLEGE LIBRARIES

苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）

Science. 1991 Jun 21;252(5013):1651-6.

Complementary DNA sequencing: expressed sequence tags and human genome project.

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al.

Source

Section of Receptor Biochemistry and Molecular Biology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD.

Abstract

Automated partial DNA sequencing was conducted on more than **600** randomly selected human brain complementary DNA (cDNA) clones to generate expressed sequence tags (ESTs).

ESTs have applications in the **discovery of new human genes**, **mapping of the human genome**, and **identification of coding regions** in genomic sequences.

Of the sequences generated, **337 represent new genes**, including 48 with significant similarity to genes from other organisms, such as a yeast RNA polymerase II subunit; Drosophila kinesin, Notch, and Enhancer of split; and a murine tyrosine kinase receptor. **Forty-six ESTs were mapped to chromosomes after amplification by the polymerase chain reaction.**

This fast approach to cDNA characterization will facilitate the tagging of most human genes in a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, and serve as a resource in diverse biological research fields.

PMID: 2047873



Nature. 1992 Feb 13;355(6361):632-4.

Sequence identification of 2,375 human brain genes.

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC.

Source

Receptor Biochemistry and Molecular Biology Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892.

Abstract

We recently described a new approach for the rapid characterization of expressed genes by partial DNA sequencing to generate 'expressed sequence tags'.

From a set of **600** human brain complementary DNA clones, **348** were informative nuclear-encoded messenger RNAs.

We have now partially sequenced **2,672** new, independent cDNA clones isolated from four human brain cDNA libraries to generate **2,375** expressed sequence tags to nuclear-encoded genes.

These sequences, together with **348** brain expressed sequence tags from our previous study, comprise more than **2,500** new human genes and **870,769** base pairs of DNA sequence.

These data represent an approximate doubling of the number of human genes identified by DNA sequencing and may represent as many as 5% of the genes in the human genome.

PMID: 1538749 work together in the cell as protein machines.

URL: DOEGenomesToLife.org

10/02



GENOMES to LIFE

BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES



如何找到EST database?

Genes and other DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY PROTEIN MACHINES

CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

goal
CHARACTERIZE GENE REGULATORY NETWORKS

WORKING CELL

Many protein machines interact through complex, interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: DOEGenomesToLife.org

10/02



NCBI >> Genbank

NCBI Resources How To

Sign in to NCBI

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

The National Biotechnology Information Center provides science and health by

About the NCBI | Mission | Organization | Research | NCBI News

DNA SEQUENCE DATA FROM GENOME PROJECTS

COMMUNITY OF CELLS

goal DEVELOP TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

IDENTIFY PROTEIN

3D Structures

Explore three-dimensional structures of proteins, DNA, and RNA molecules. Examine sequence-structure relationships, active sites, molecular interactions, biological activities of bound chemicals, and associated biosystems.

II 1 2 3 4 5 6 7 8

goal CHARACTERIZE GENE REGULATORY NETWORKS

Many protein machines interact through complex interconnected processes. These dynamic processes will lead to Nov 26, 2014

NCBI Announcements

NCBI to hold two-day genomics hackathon in January

From January 5th to 7th, NCBI will host a genomic hackathon focusing on

NCBI BioSample includes curated list of over 400 known misidentified and contaminated cell lines

Nov 24, 2014

点此打开数据库选择列表

A red box highlights the 'All Databases' dropdown menu in the top navigation bar. A red arrow points from the text '点此打开数据库选择列表' (Click here to open the database selection list) to this highlighted menu item.



NCBI >> Genbank

NCBI Resources How To

Apply knowledge of

Sign in to NCBI



National Center for
Biotechnology Information

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

Proteins perform many of the most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Explore three-dimensional structures of proteins, DNA, and RNA molecules. Examine sequence-structure relationships, active sites, molecular interactions, biological activities of bound chemicals, and associated biosystems.

1 2 3 4 5 6 7 8

- All Databases
- Recent
- All Databases
- PubMed
- All
- All Databases
- Assembly
- BioProject
- BioSample
- BioSystems
- Books
- ClinVar
- Clone
- Conserved Domains
- dbGaP
- dbVar
- Epigenomics
- EST**
- Gene
- Genome
- GEO DataSets

点击这里
选择EST数据库

NCBI
LIVE APPROACHES
Center for Biotechnology Information advances science and health by
to biomedical and genomic information.

NCBI | Mission | Organization | Research | NCBI News

analyze data using NCBI software
s: Get NCBI data or software
Learn how to accomplish specific tasks at NCBI
ons: Submit data to GenBank or other NCBI databases

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

NCBI to hold two-day genomics
hackathon in January

From January 5th to 7th, NCBI will host a
genomics hackathon focusing on

NCBI BioSample includes curated list of
over 400 known misidentified and
contaminated cell lines

Nov 24, 2014



NCBI >> Genbank



NCBI >> Genbank >> EST Homepage

NCBI Resources How To

Sign in to NCBI

EST

EST

Limits Advanced

Search

Help

EST

The EST database is a collection of short single-read transcript sequences from GenBank. These sequences provide a resource to evaluate gene expression, find potential variation, and annotate genes.

Using EST

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

EST Tools

[Submit to EST](#)

[BLAST](#)

[LinkOut](#)

[E-Utilities](#)

[Batch Entrez](#)

Other Resources

[GenBank Home](#)

[UniGene](#)

[SRA Home](#)

[INSDC](#)

PROTEINS

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: [DOEGenomesToLife.org](#)

10/02



NCBI >> Genbank

NCBI Resources How To

Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases

Nucleotide
OMIM
PMC
PopSet
Probe
Protein
Protein Clusters
PubChem BioAssay
PubChem Compound
PubChem Substance
PubMed
PubMed Health
SNP
SRA
Structure
Taxonomy
ToolKit
ToolKitAll
ToolKitBook
UniGene

Popular Resources

SNP
Gene
Protein
PubChem

NCBI

National Center for Biotechnology Information to biomedical and genomic information

Mission | Organization | Resources

Analyze data using NCBI software
Get NCBI data or software

Learn how to accomplish specific tasks at NCBI

Submit data to GenBank or other NCBI databases

选择UniGene数据库后
使用同样的方法进入
UniGene数据库主页

NCBI YouTube channel

Learn how to get the most out of NCBI tools and databases with video tutorials on the NCBI YouTube Channel.

GO

II 1 2 3 4 5 6 7 8

NCBI Announcements

NCBI to hold two-day genomics hackathon in January

Nov 26, 2014

From January 5th to 7th, NCBI will host a genomic hackathon focusing on

NCBI BioSample includes curated list of over 400 known misidentified and contaminated cell lines

Nov 24, 2014



NCBI >> Genbank >> UniGene Homepage

NCBI Resources How To

UniGene

UniGene

Limits Advanced Help

UniGene

UniGene Tools

[UniGene Statistics](#)

[Library browser](#)

[Digital Differential Display \(DDD\)](#)

[Downloads/FTP](#)

UniGene

UniGene computationally identifies transcripts from the same locus; analyzes expression by tissue, age, and health status; and reports related proteins (protEST) and clone resources.

goal
DEVELOP
COMPUTATION
CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

UniGene数据库统计数据
按物种进行分类统计

Using UniGene

Getting Started

FAQ

Query Tips

ProtEST

DDD

Clustering by transcripts

Clustering by genomes

many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

Gene

HomoloGene

Map Viewer

Short Read Archive (SRA)

URL: [DOEGenomesToLife.org](#)

10/02



NCBI >> Genbank >> UniGene >> Statistics

NCBI Resources How To

Apply knowledge of

Sign in to NCBI

UniGene

UniGene

Limits Advanced

Search

Help

UniGene Statistics

| Species | UniGene entries | Cutoff Date* | Release Date* |
|---|------------------------|--------------|---------------|
| Chordata | | | |
| Mammalia | | | |
| Bos taurus (cow) | 45364 | Jun 19 2012 | Jul 9 2012 |
| Canis lupus familiaris (dog) | 23847 | Sep 22 2012 | Jan 2 2013 |
| Capra hircus (domestic goat) | 38271 | Nov 29 2011 | Dec 20 2011 |
| Equus caballus (horse) | 6289 | Mar 17 2011 | May 7 2012 |
| Homo sapiens (human) | 130029 | Sep 22 2012 | Nov 13 2012 |
| Macaca fascicularis (crab-eating macaque) | 12546 | Aug 28 2011 | Nov 7 2011 |
| Macaca mulatta (rhesus monkey) | 26849 | Aug 5 2012 | Sep 10 2012 |
| Monodelphis domestica (gray short-tailed opossum) | 662 | Oct 24 2010 | Jan 31 2012 |
| Mus musculus (mouse) | 80383 | Jun 20 2012 | Jul 11 2012 |
| Ornithorhynchus anatinus (platypus) | 749 | Nov 26 2009 | Jul 27 2010 |
| Oryctolagus cuniculus (rabbit) | 7298 | Mar 25 2012 | Apr 3 2012 |
| Ovis aries (sheep) | 17318 | Aug 27 2011 | Nov 21 2011 |
| Pan troglodytes (chimpanzee) | 2412 | Mar 15 2011 | Mar 28 2011 |
| Papio anubis (olive baboon) | 11659 | Mar 8 2011 | Nov 7 2011 |
| Peromyscus maniculatus (deer mouse) | 11045 | Mar 9 2010 | Aug 3 2011 |



NCBI >> Genbank >> UniGene

NCBI Resources How To

Protect workers
Apply knowledge of
microbial function

Sign in to NCBI

UniGene

UniGene

BIOLOGICAL SOLUTIONS
DNA PROFILING

environment

Search

Limits Advanced

Help



UniGene

UniGene computationally identifies transcripts from the same locus; analyzes expression by tissue, age, and health status; and reports related proteins (protEST) and clone resources.

Using UniGene

Genes and other
DNA sequences
Contain instructions
on how and when
to build proteins

Query Tips

ProtEST

DDD

Clustering by transcripts

Clustering by genomes · of life's most essential functions. To carry out their
specific roles, they often work together in the cell as protein machines.

UniGene Tools

[UniGene Statistics](#)

[Library browser](#)

[Digital Differential Display \(DDD\)](#)

[Downloads/FTP](#)

Also of interest

EST文库浏览

[Gene](#)

[HomoloGene](#)

[Map Viewer](#)

[Short Read Archive \(SRA\)](#)

machines interact
through complex
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life

URL: DOEGenomesToLife.org

10/02



NCBI >> Genbank >> UniGene >> Library Browser

Search UniGene

Libraries for Homo sapiens with minimum sequences 1000

Body Sites

adipose tissue

| Lib. ID | Library Name | Sequences |
|-----------|---|-----------|
| Lib.10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 9638 |
| Lib.886 | NCI_CGAP_Lip2 | 1740 |
| Lib.16445 | Sugano cDNA library, adipose tissue | 1665 |
| Lib.816 | Adipose tissue, white II | 1195 |

Not Shown: 14 libraries having fewer than 1000 sequences

adrenal gland

| Lib. ID | Library Name | Sequences |
|-----------|--|-----------|
| Lib.18302 | ADRGL2 | 10385 |
| Lib.7317 | NIH_MGC_84 | 7572 |
| Lib.6791 | ADB | 6475 |
| Lib.927 | NCI_CGAP_AA1 | 3363 |
| Lib.16377 | Sugano cDNA library, adrenal gland | 2772 |
| Lib.6815 | Protein machines interact in many of these dynamic processes. will lead to many new insights into how protein machines carry out their often complex functions. They often carry out their functions together in the cell as protein machines. | 2460 |
| Lib.6792 | ADC | 1995 |
| Lib.6793 | Cu | 1649 |
| Lib.993 | NCI_CGAP_Phe1 | 1356 |

EST文库ID号

苏州大学 医学院 基础医学与生物科学学院 生物信息教研室 张高川 (未经同意不得私自转载和发布) 1183

NCBI >> Genbank >> UniGene >> Library Browser

NCBI ◊ UniGene ◊ Homo sapiens ◊ Library Browser

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search UniGene

Libraries for Homo sapiens with minimum sequences 1000

Produce and use energy
goal EXPLORE FUNCTION IN MICROBIAL COMMUNITIES

Collapse All | Expand All

Body Sites

adipose tissue

| Lib. ID | Library Name | Sequences |
|-----------|---|-----------|
| Lib.10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 9638 |
| Lib.886 | NCI_CGAP_Lip2 | 1740 |
| Lib.16445 | Sugano cDNA library, adipose tissue | 1665 |
| Lib.816 | Adipose tissue, white II | 1195 |

fewer than 1000 sequences

EST文库名称

adrenal gland

| Lib. ID | Library Name | Sequences |
|-----------|------------------------------------|-----------|
| Lib.18302 | ADRGL2 | 10385 |
| Lib.7317 | NIH_MGC_84 | 7572 |
| Lib.6791 | ADB | 6475 |
| Lib.927 | NCI_CGAP_AA1 | 3363 |
| Lib.16377 | Sugano cDNA library, adrenal gland | 2772 |
| Lib.6815 | cdA | 2460 |
| Lib.6792 | ADC | 1995 |
| Lib.6793 | Cu | 1002 |
| Lib.993 | NCI_CGAP_Phe1 | 1649 |

30 libraries

protein machines interact in complex pathways to carry out their dynamic processes, will lead to many of life's processes.

CHARACTERIZE GENE REGULATORY NETWORKS

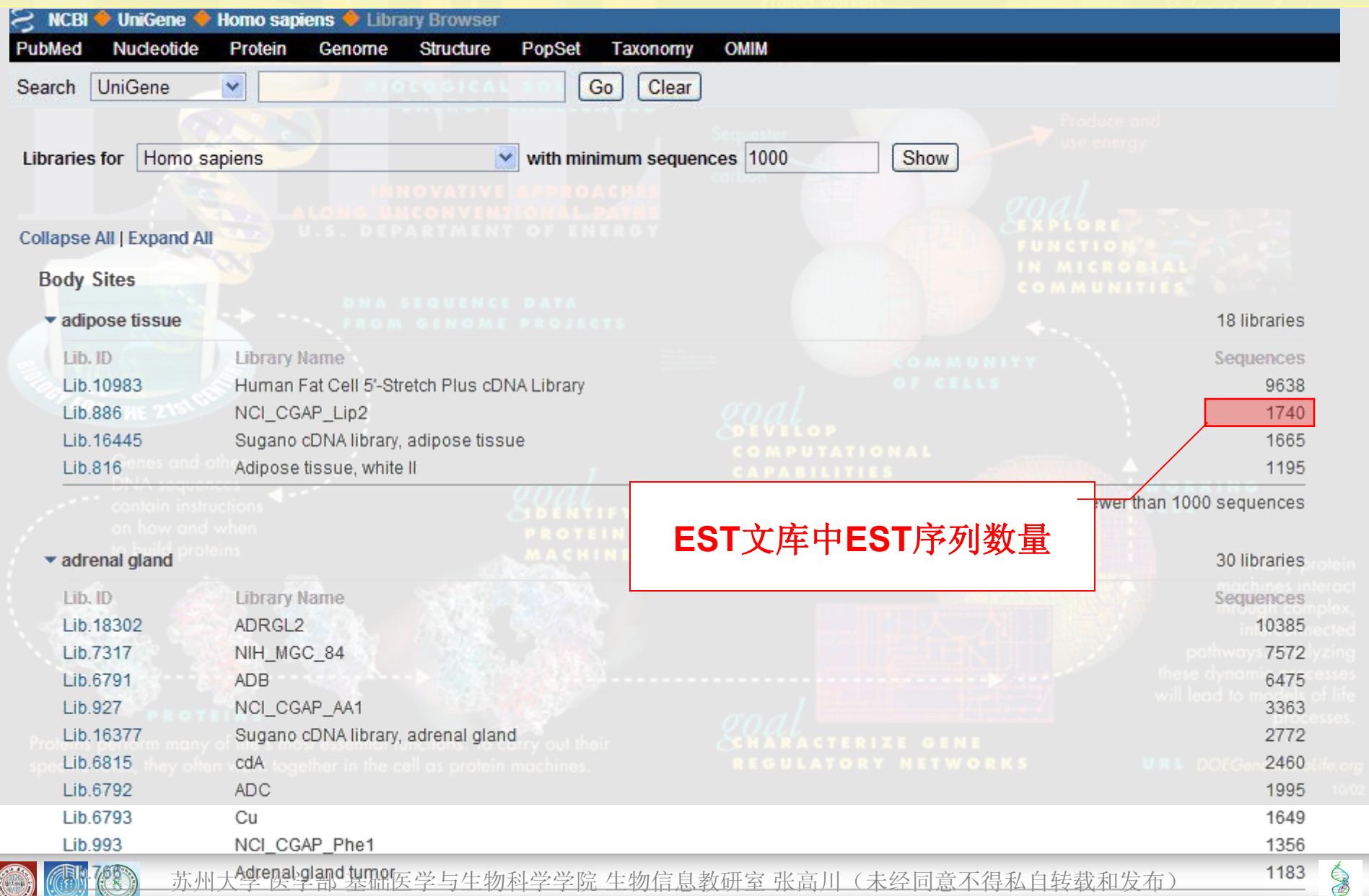
URL: DOEGenome.org

1183

苏州大学 医学院 基础医学与生物科学学院 生物信息教研室 张高川 (未经同意不得私自转载和发布)

Adrenal-gland tumor

NCBI >> Genbank >> UniGene >> Library Browser



NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*
Tissue: liposarcoma
Vector: pAMP10
Vector type: plasmid (ampicillin resistant)
Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Genes and other DNA

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [#236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate level in transcripts per million (TPM).

| ESTs | TPM | | UniGene Entry |
|------|-------|---|--|
| 30 | 45386 | ● | Hs.510635 Immunoglobulin heavy constant gamma 1 (G1m marker). |
| 24 | 36309 | ● | Hs.449585 Immunoglobulin lambda constant 1 (Mcg marker). |
| 23 | 34796 | ● | Hs.449609 Immunoglobulin kappa constant. |
| 6 | 9077 | ● | Hs.74034 Caveolin 1, caveolae protein, 22kDa. Protein score: 3.000. This marker is specific for their |
| 6 | 9077 | ● | Hs.275243 S100 calcium binding protein A6. |
| 4 | 6051 | ● | Hs.8102 Ribosomal protein S20. |
| 4 | 6051 | ● | Hs.744846 Ribosomal protein S14. |

Clean up the environment

Sequester

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

EST文库描述信息

物种来源

组织来源

质粒载体

宿主细胞

作者及其参考文献



NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT prim selected on agarose gel, average insert size 600 bp. Refe Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Genes and other DNA

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | | UniGene Entry |
|------|-------|---|---|
| 30 | 45386 | ● | Hs.510635 Immunoglobulin heavy constant gamma 1 (G1m marker). |
| 24 | 36309 | ● | Hs.449585 Immunoglobulin lambda constant 1 (Mcg marker). |
| 23 | 34796 | ● | Hs.449609 Immunoglobulin kappa constant. |
| 6 | 9077 | ● | Hs.74034 Caveolin 1, caveolae protein, 22kDa. |
| 6 | 9077 | ● | Hs.275243 S100 calcium binding protein A6. |
| 4 | 6051 | ● | Hs.8102 Ribosomal protein S20. |
| 4 | 6051 | ● | Hs.744846 Ribosomal protein S14. |

EST文库数量
提交时间

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- Homepage
- News
- Help
- Library Browser
- DDD
- Download UniGene

Related Databases

- Gene
- Gene Expression



NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT priming. No selected on agarose gel, average insert size 600 bp. Reference: Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Genes and other
DNA

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | | UniGene Entry |
|------|-------|---|---|
| 30 | 45386 | ● | Hs.510635 Immunoglobulin heavy constant gamma 1 (G1m marker). |
| 24 | 36309 | ● | Hs.449585 Immunoglobulin lambda constant 1 (Mcg marker). |
| 23 | 34796 | ● | Hs.449609 Immunoglobulin kappa constant. |
| 6 | 9077 | ● | Hs.74034 Caveolin 1, caveolae protein, 22kDa. |
| 6 | 9077 | ● | Hs.275243 S100 calcium binding protein A6. |
| 4 | 6051 | ● | Hs.8102 Ribosomal protein S20. |
| 4 | 6051 | ● | Hs.744846 Ribosomal protein S14. |

点击该按钮
可以下载整个文库

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

[Download Sequences](#)

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- [Homepage](#)
- [News](#)
- [Help](#)
- [Library Browser](#)
- [DDD](#)
- [Download UniGene](#)

Related Databases

- [Gene](#)
- [Gene Expression](#)



NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Genes and other
DNA

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be viewed in the table below. TPM = transcript per million level in transcripts per million (TPM).

| ESTs | TPM | | |
|------|-------|-------------|---------------------------------|
| 30 | 45386 | ● Hs.510635 | Immunoglobulin heavy chain |
| 24 | 36309 | ● Hs.449585 | Immunoglobulin lambda chain |
| 23 | 34796 | ● Hs.449609 | Immunoglobulin kappa chain |
| 6 | 9077 | ● Hs.74034 | Caveolin 1, caveolae protein |
| 6 | 9077 | ● Hs.275243 | S100 calcium binding protein A1 |
| 4 | 6051 | ● Hs.8102 | Ribosomal protein S20 |
| 4 | 6051 | ● Hs.744846 | Ribosomal protein S14 |

EST文库关键词
组织细胞类型
病理类型
发育阶段
等信息

Clean up the environment

Sequester excess

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- Homepage
- News
- Help
- Library Browser
- DDD
- Download UniGene

Related Databases

- Gene
- Gene Expression



NCBI >> Genbank >> UniGene >> Library Browser >> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | | UniGene Entry |
|------|-------|---|---|
| 30 | 45386 | ● | Hs.510635 Immunoglobulin heavy constant gamma 1 (G1m marker). |
| 24 | 36309 | ● | Hs.449585 Immunoglobulin lambda constant 1 (Mcg marker). |
| 23 | 34796 | ● | Hs.449609 Immunoglobulin kappa constant. |
| 6 | 9077 | ● | Hs.74034 Caveolin 1, caveolae protein, 22kDa. |
| 6 | 9077 | ● | Hs.275243 S100 calcium binding protein A6. |
| 4 | 6051 | ● | Hs.8102 Ribosomal protein S20. |
| 4 | 6051 | ● | Hs.744846 Ribosomal protein S14. |

EST文库分析结果

与已知UniGene匹配的EST数量

基于EST片段计数的基因表达水平的标准化

EST Sequences

1,740 EST sequences from this library have

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- Homepage
- News
- Help
- Library Browser
- DDD
- Download UniGene

Related Databases

- Gene
- Gene Expression

NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

Genes and other
DNA

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|-----|--|
| 30 | | Ig-globulin heavy constant gamma 1 (G1m marker). |
| 24 | | Ig-globulin lambda constant 1 (Mcg marker). |
| 23 | | Ig-globulin kappa constant. |
| 6 | | Ann 1, caveolin-1, rotavirus 2kDa, their |
| 6 | | calcium binding protein A6. |
| 4 | | 1 |
| 4 | | 11, human Kappa light chain protein S14. |

661 与 504 之
间的数量差异意
味着什么？

Clean up the
environment

Sequester
exon

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- Homepage
- News
- Help
- Library Browser
- DDD
- Download UniGene

Related Databases

- Gene
- Gene Expression



NCBI >> Genbank >> UniGene >> Library Browser

>> NCI_CGAP_Lip2

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:886 (dbEST ID)

NCI_CGAP_Lip2

Library Description

Organism: *Homo sapiens*

Tissue: liposarcoma

Vector: pAMP10

Vector type: plasmid (ampicillin resistant)

Lab host: DH10B

mRNA made from liposarcoma, cDNA made by selected on agarose gel, average insert size Research 56:5380-5383.

Source: Robert Strausberg, Ph.D.

661 与 1740 之间的数量差异又意味着什么？

Gene Content Analysis

661 ESTs from this library were grouped into 504 UniGene entries (negative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | | UniGene Entry |
|------|-------|---|---|
| 30 | 45386 | ● | Hs.510635 Immunoglobulin heavy constant gamma 1 chain marker. |
| 24 | 36309 | ● | Hs.449585 Immunoglobulin lambda chain marker. |
| 23 | 34796 | ● | Hs.449609 Immunoglobulin kappa constant chain marker. |
| 6 | 9077 | ● | Hs.74034 Caveolin 1, caveolae protein, 22kDa. |
| 6 | 9077 | ● | Hs.275243 S100 calcium binding protein A6. |
| 4 | 6051 | ● | Hs.8102 Ribosomal protein S20. |
| 4 | 6051 | ● | Hs.744846 Ribosomal protein S14. |

Clean up the environment
Sequester excess

EST Sequences

1,740 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Keywords

- adipose tissue
- soft tissue/muscle tissue tumor
- unknown developmental stage
- non-normalized

Source: NCBI Clone Registry

UniGene Links

- Homepage
- News
- Help
- Library Browser
- DDD
- Download UniGene

Related Databases

- Gene
- Gene Expression



NCBI >> Genbank >> UniGene >> human Cytochrome C

Apply knowledge of

NCBI Resources How To Sign in to NCBI

UniGene UniGene human cytochrome C Search Help

Save search Limits Advanced

Display Settings: Summary, 20 per page Send to: Filter your results:

Results: 1 to 20 of 114 << First < Prev Page 1 of 6 Next > Last >>

[Cytochrome P450, family 2, subfamily C, polypeptide 9](#)
1. **CYP2C9, Homo sapiens**
Hs.282624: 202 sequences.
[Order cDNA clone](#)

[Cytochrome P450, family 2, subfamily C, polypeptide 19](#)
2. **CYP2C19, Homo sapiens**
Hs.282409: 9 sequences.

[Immunoglobulin heavy constant mu](#)
3. **Ighm, Mus musculus**
Mm.342177: 1687 sequences.
[Order cDNA clone](#)

[Immunoglobulin heavy chain \(V3660 family\)](#)
4. **Igh-V3660, Mus musculus**
Mm.474964: 190 sequences.

All (114)
Fungi (0)
Insects (0)
[Mammals \(114\)](#)
Plants (0)

[Manage Filters](#)

▼ Top Organisms [Tree]
Homo sapiens (88)
Macaca fascicularis (15)
Mus musculus (7)
Canis lupus familiaris (3)
Bos taurus (1)

Find related data
Database: Select

Find items

 苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川 (未经同意不得私自转载和发布)

NCBI >> Genbank >> UniGene >> human Cytochrome P450

Apply knowledge of

NCBI

UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

UGID:176250 UniGene Hs.282624 *Homo sapiens* (human) CYP2C9 Order cDNA clone, Links

Cytochrome P450, family 2, subfamily C, polypeptide 9 (CYP2C9)

Human protein-coding gene CYP2C9. Represented by 189 ESTs from 27 cDNA libraries. EST representation biased toward adult. Corresponds to reference sequence NM_000771.3. [UniGene 176250 - Hs.282624]

SELECTED PROTEIN SIMILARITIES

Comparison of cluster transcripts with RefSeq proteins. The alignments can suggest function of the cluster.

| | Best Hits and Hits from model organisms | Species | Id(%) | Len(aa) |
|---|--|-----------------------|-------|---------|
| NP_000762.2 | CYP2C9 gene product | <i>H. sapiens</i> | 100.0 | 489 |
| XP_003339236.1 | PREDICTED: cytochrome P450 2C9-like isoform 1 | <i>P. troglodytes</i> | 99.6 | 489 |
| NP_031841.3 | Cyp2c29 gene product | <i>M. musculus</i> | 81.0 | 489 |
| NP_001085245.1 | cytochrome P450, family 2, subfamily C, polypeptide 18 | <i>X. laevis</i> | 58.6 | 486 |
| XP_001923597.2 | PREDICTED: cytochrome P450 2G1-like | <i>D. rerio</i> | 52.5 | 488 |
| Other hits (2 of 28) [Show all] | | Species | | |
| NP_001035301.1 | CYP2C75 gene product | <i>M. mulatta</i> | | |

向下翻页



NCBI >> Genbank >> UniGene >> human Cytochrome P450 >> mRNA

SEQUENCES

Sequences representing this gene; mRNAs, ESTs, and gene predictions supported by transcribed sequences.

mRNA sequences (13)

- | | | |
|-----------------------------|--|----------|
| BC070317.1 | Homo sapiens cytochrome P450, family 2, subfamily C, polypeptide 9, PA mRNA (cDNA clone IMAGE:4722940), complete cds | |
| M15331.1 | Human liver cytochrome P-450 S-mephentyoin 4-hydroxylase (P- 450mp) mRNA, partial cds | P |
| M21940.1 | Human cytochrome P-450 S-mephentyoin 4-hydroxylase (P-450mp) mRNA, 3' end, clone MP-4 | P |
| D00173.1 | Homo sapiens mRNA for cytochrome P-450, partial cds | P |
| NM_000771.3 | Homo sapiens cytochrome P450, family 2, subfamily C, polypeptide 9 PA (CYP2C9), mRNA | |
| BC125054.1 | Homo sapiens cytochrome P450, family 2, subfamily C, polypeptide 9, P mRNA (cDNA clone MGC:149605 IMAGE:40116637), complete cds | |
| BC020754.1 | Homo sapiens cytochrome P450, family 2, subfamily C, polypeptide 9, PA mRNA (cDNA clone IMAGE:4766890), complete cds | |
| AK289420.1 | Homo sapiens cDNA FLJ75561 complete cds, highly similar to Homo PA sapiens cytochrome P450, family 2, subfamily C, polypeptide 9 (CYP2C9), mRNA | |
| AK298458.1 | Homo sapiens cDNA FLJ58095 complete cds, highly similar to P Cytochrome P450 2C9 (EC 1.14.13.80) | |
| M21939.1 | Human cytochrome P-450 S-mephentyoin 4-hydroxylase (P-450mp) | |

向下翻页



NCBI >> Genbank >> UniGene >> human Cytochrome P450 >> EST

EST sequences (10 of 189) [Show all sequences]

| | | | |
|------------|--|-------|-----------|
| BX095470.1 | Clone IMAGp998P17372_:_IMAGE:195712 | mixed | |
| CB148958.1 | Clone L17N670205-1-A06 | liver | 5' read P |
| CB148984.1 | Clone L17N670205-1-E03 | liver | 5' read P |
| CB148999.1 | Clone L17N670205-1-G09 | liver | 5' read P |
| CB154328.1 | Clone L17N670205-10-B06 | liver | 5' read P |
| CB161229.1 | Clone L17N670205n1-18-F12 | liver | 5' read P |
| CB161481.1 | Clone L17N670205n1-34-C02 | liver | 5' read A |
| CB161916.1 | Clone L17N670205n1-41-H06 | liver | 5' read P |
| CB162071.1 | Clone L17N670205n1-16-H08 | liver | 5' read P |
| CB162668.1 | Clone L17N670205n1-45-H01 | liver | 5' read P |

[Download Sequences](#)

Key to Symbols

- P Has similarity to known Proteins (after translation)
- A Contains a poly-Adenylation signal
- S Sequence is a Suboptimal member of this cluster
- M Clone is putatively CDS-complete by MGC criteria

NCBI >> Genbank >> UniGene >> human Cytochrome P450 >> EST sequences

>gnl|UG|Hs#S11124513 BX095470 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGP998P17372, mRNA sequence /clone=IMAGP998P17372_ ; IMAGE:195712

/gb=BX095470 /gi=27827717 /ug=Hs.282624 /len=601

```
CTGGATGAAGGTGGCAATTAAAGAAAAAGTAAATACTCATGCCCTTCTCAGCAGGAAAA  
CGGATTGTGTGGAGAAGCCCTGGCCGGCATGGAGCTTTTATTCCCTGACCTCCATT  
TTACAGAACTTTAACCTGAAATCTCTGGTGACCCAAAGAACCTTGACACCACTCCAGTT  
GTCAATGGATTCCCTGTGCCGCCCTCTACAGCTGTGCTTCATTCCGTCTGAAGAAG  
AGCAGATGCCCTGGCTGCTGCTGCAGTCCTGCAGCTCTTTCCCTGGGGCATTAT  
CCATCTTCACATCTGTAATGCCCTTCTCACCTGTACATCTCACATTTCCTCCCTG  
AAGATCTAGTGAACATTGACCTCCATTACGGAGAGTTCCATGTTCACTGTGCAAAT  
ATATCTGCTATTCTCCATACTCTGTAACAGTTGCATTGACTGTACATAATGCTCATACT  
TATCTAATGTTGAGTTATTAAATATGTTATTAAATAGAGAAATATGATTGTGATTAA  
TAATTCAAAGGCATTCTTCTGCATGTTCTAAATAAAAAGCATTATTATTGCTGAAN  
A
```

>gnl|UG|Hs#S11531425 K-EST0205381 L17N670205 Homo sapiens cDNA clone L17N670205-1-A06 5', mRNA sequence /clone=L17N670205-1-A06 /clone_end=5'

/gb=CB148958 /gi=28131667 /ug=Hs.282624 /len=466

```
AGTAAAAGAACACCAAGAACATGGACATGAACAAACCTCAGGACTTTATTGATTGCTT  
CCTGATGAAAATGGAGAACGCTGGAAAAACACTGCAGTTGACTTGTGGAGCTGGGACAG  
AGACGACAAGCACAAACCTGAGATATGCTCTCTCTGCTGAAGCACCCAGAGGTCA  
CAGCTAAAGTCCAGGAAGAGATTGAACGTGTGATTGGAGAAACCGGAGCCCTGCATGC  
AAGACAGGGAGCCACATGCCCTACACAGATGCTGTGGTGACGAGGTCCAGAGATACTTG  
ACCTTCTCCCCACCAGCCTGCCCTATGCAGTGACATTAAATTCAAGAAACTATC  
TCATTCCCAAGGGACAACCATATTAAATTCCCTGACTTCTGTGCTACATGACAACAAAG  
AATTTCCTAACCCAGAGATGTTGACCCCTCATCACTTCTGGATGA--
```

>gnl|UG|Hs#S11531439 K-EST0205426 L17N670205 Homo sapiens cDNA clone

L17N670205-1-E03 5', mRNA sequence /clone=L17N670205-1-E03 /clone_end=5'

/gb=CB148984 /gi=28131735 /ug=Hs.282624 /len=524

```
GAAAAAGTTATATTGGAAAAAGTAAAAGAACACCAAGAACATGGACATGAACAAACCC  
TCAGGACTTTATTGATTGCTTCCTGATGAAAATGGAGAACGGAAAAGCACAAACCAACCATC  
TGAATTACTATTGAAAGCTGGAAAAACACTGCAGTTGACTTGTGGAGCTGGGACAGA  
GACGCACAAAGCACACCCAGAGATATGCTCTCTCTGCTGAAGCACCCAGAGGTCAAC  
AGCTAAAGTCCAGGAAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCATGCA
```

FASTA format

Apply knowledge of
biological functional
abilities

WORKING
CELL

Many protein
machines interact
through complex
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life
processes.

URL: DOEGenomesToLife.org

10/02

Proteins perform
specific roles



苏州大学 医学院 基础医学与生物科学学院 生物信息教研室 张高川 (未经授权不得私自转载和发布)



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

NCBI Resources How To Sign in to NCBI

UniGene UniGene Search Limits Advanced Help

UniGene

UniGene computationally identifies transcripts from the same locus; analyzes expression by tissue, age, and health status; and reports related proteins (protEST) and clone resources.

Using UniGene

- Getting Started
- FAQ
- Query Tips
- ProtEST
- DDD
- Clustering by transcripts
- Clustering by genomes

UniGene Tools

- UniGene Statistics
- Library browser
- Digital Differential Display (DDD)
- Downloads/FTP

Also of interest

- BioSample
- EST sequences
- Gene
- HomoloGene
- Map Viewer
- Short Read Archive (SRA)

You are here: NCBI > Genes & Expression > UniGene

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Support Center

DOEGenomesToLife.org

10/02

数字化示差分析法



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

NCBI UniGene ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene

Homepage FAQs Help Query Tips Library Browser DDD Download UniGene

Related Databases Gene HomoloGene EST SRA

NIH cDNA Projects

NLM | NIH | UniGene | Privacy Statement | Disclaimers

选择某一个物种进行分析,
这里以人类 (human) 为例

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

NCBI UniGene ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human)

Pool A: Normal DCs

Start Over Continue

Define a cDNA 'pool' by selecting one or more cDNA libraries from the list below. Also enter a short descriptive label for the pool in the text field above.

| Lib.ID | cDNA Library Name | Clustered ESTs |
|--------------------------------|---|----------------|
| Adipose tissue | | |
| <input type="checkbox"/> 814 | Adipose tissue, white I | 335 |
| <input type="checkbox"/> 816 | Adipose tissue, white II | 976 |
| <input type="checkbox"/> 886 | NCI_CGAP_Lip2 | 661 |
| <input type="checkbox"/> 10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 8140 |
| <input type="checkbox"/> 16445 | Sugano cDNA library, adipose tissue | 1641 |
| <input type="checkbox"/> 18329 | ADIPS2 | 460 |
| Adrenal gland | | |
| <input type="checkbox"/> 766 | Adrenal gland tumor | 962 |

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)

10/02

A组为正常的分化细胞
(Normal Differentiated Cells)



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human) Start Over

Pool A: Normal DCs

Define a cDNA 'pool' by selecting one

LIBID

Adipose tissue

- 814 Adipose tissue, white I 335
- 816 Adipose tissue, white II 976
- 886 NCI_CGAP_Lip2 661
- 10983 Human Fat Cell 5'-Stretch Plus cDNA Library 8140
- 16445 Sugano cDNA library, adipose tissue 1641
- 18329 ADIPS2 460

Adrenal gland

- 766 Adrenal gland tumor 962

Related Databases Clustered ESTs

NIH cDNA Projects Finding cDNAs

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)

10/02

这是复选框，勾中可以选择相应的EST/cDNA library



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human)

Pool A: Normal DCs

Start Over Continue

Define a cDNA 'pool' by selecting one or more cDNA libraries.

Lib.ID

Adipose tissue

| Lib.ID | Sample Description | Number of ESTs |
|--------|---|----------------|
| 814 | Adipose tissue, white I | 335 |
| 816 | Adipose tissue, white II | 976 |
| 886 | NCI_CGAP_Lip2 | 661 |
| 10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 8140 |
| 16445 | Sugano cDNA library, adipose tissue | 1641 |
| 18329 | ADIPS2 | 460 |

Adrenal gland

| Lib.ID | Sample Description | Number of ESTs |
|--------|---------------------|----------------|
| 766 | Adrenal gland tumor | 962 |

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)

10/02



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

NCBI

UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene Homepage FAQs Help Query Tips Library Browser DDD Download UniGene

Related Databases Gene HomoloGene EST SRA

NIH cDNA Projects Finding cDNAs

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human) Start Over

Pool A: Normal DCs

Define a cDNA 'pool' by selecting one or more cDNA libraries from the list below.

| Lib.ID | | ESTs |
|--------|---|------|
| 814 | Adipose tissue, white I | 335 |
| 816 | Adipose tissue, white II | 976 |
| 886 | NCI_CGAP_Lip2 | 661 |
| 10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 8140 |
| 16445 | Sugano cDNA library, adipose tissue | 1641 |
| 18329 | ADIPS2 | 460 |
| 766 | Adrenal gland tumor | 962 |

这是EST/cDNA libraries的名称，
有时名称太简单，无法区分样本
具体的病理、生理分类

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)

10/02



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

Apply knowledge of

NCBI UniGene ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human) Start Over

Pool A: Normal DCs

Define a cDNA 'pool' by selecting one or more cDNA libraries in the list below. You can also type the library name in the search field above.

Lib.ID

Adipose tissue

- 814 Adipose tissue, white I
- 816 Adipose tissue, white II
- 886 NCI_CGAP_Lip2
- 10983 Human Fat Cell 5'-Stretch Plus cDNA Library
- 16445 Sugano cDNA library, adipose tissue
- 18329 ADIPS2

Adrenal gland

- 766 Adrenal gland tumor

Clustered ESTs

| |
|------|
| 335 |
| 976 |
| 661 |
| 8140 |
| 1641 |
| 460 |
| 962 |

NIH cDNA Projects Finding cDNAs

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)

10/02

这是EST/cDNA libraries 中的EST序列数目



NCBI >> Genbank >> UniGene >> Digital Differential Display (DDD)

NCBI

UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene Homepage FAQs Help Query Tips Library Browser DDD Download UniGene

Related Databases Gene HomoloGene EST SRA

NIH cDNA Projects Finding cDNAs

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens (human) Start Over

Pool A: Normal DCs Continue

Define a cDNA 'pool' by selecting one or more cDNA libraries from the list below. Also enter a short descriptive label for the pool in the text field above.

| Lib.ID | cDNA Library Name | Clustered ESTs |
|--------------------------------|---|----------------|
| Adipose tissue | | |
| <input type="checkbox"/> 814 | Adipose tissue, white I | 335 |
| <input type="checkbox"/> 816 | Adipose tissue, white II | 976 |
| <input type="checkbox"/> 886 | NCI_CGAP_Lip2 | 661 |
| <input type="checkbox"/> 10983 | Human Fat Cell 5'-Stretch Plus cDNA Library | 8140 |
| <input type="checkbox"/> 16445 | Sugano cDNA library, adipose tissue | 1641 |
| <input type="checkbox"/> 18329 | ADIPS2 | 460 |
| Adrenal gland | | |
| <input type="checkbox"/> 766 | Adrenal gland tumor | 962 |



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Apply knowledge of





Search UniGene



for



NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:239 (dbEST ID)

Human White blood cells

Library Description

Organism: *Homo sapiens*
 Organ: blood
 Tissue: white blood cells

Source: Venter, JC

文库描述信息
(人类血液中的白细胞)

have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

Download Sequences

Click the button to save sequences to a file (FASTA format)

Gene Content Analysis

911 ESTs from this library were grouped into 707 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|------|--|
| 9 | 9879 | Hs.530096 Eukaryotic translation initiation factor 3, subunit I. |
| 9 | 9879 | Hs.524741 Actin related protein 2/3 complex, subunit 3, 21kDa. |
| 7 | 7684 | Hs.75117 Interleukin enhancer binding factor 2. |
| 7 | 7684 | Hs.643544 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide. |
| 7 | 7684 | Hs.592455 Transcribed locus, strongly similar to NP_001003447.1 rpl15 gene product [<i>Danio rerio</i>]. |
| 7 | 7684 | Hs.405590 Eukaryotic translation initiation factor 3, subunit E. |
| 7 | 7684 | Hs.12013 ATP-binding cassette, sub-family E (OABP), member 1. |
| 5 | 5488 | Hs.505735 Nascent polypeptide-associated complex alpha subunit. |
| 5 | 5488 | Hs.180877 H3 histone, family 3B (H3.3B). |

Keywords

- blood
- normal
- unknown developmental stage
- white blood cell
- non-normalized
- bulk

Source: NCBI Clone Registry

References

[Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence](#). Adams MD, et al., *Nature* 377, 3-174 (1995).

UniGene Links





NCBI



Search UniGene

for

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:239 (dbEST ID)

这里给出该文库中EST序列的
总数量以及下载通道

Human White blood cells

Library Description

Organism: *Homo sapiens*

Organ: blood

Tissue: white blood cells

Source: Venter, JC

Gene Content Analysis

911 ESTs from this library were grouped into 707 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|------|--|
| 9 | 9879 | Hs.530096 Eukaryotic translation initiation factor 3, subunit I. |
| 9 | 9879 | Hs.524741 Actin related protein 2/3 complex, subunit 3, 21kDa. |
| 7 | 7684 | Hs.75117 Interleukin enhancer binding factor 2. |
| 7 | 7684 | Hs.643544 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide. |
| 7 | 7684 | Hs.592455 Transcribed locus, strongly similar to NP_001003447.1 rpl15 gene product [<i>Danio rerio</i>]. |
| 7 | 7684 | Hs.405590 Eukaryotic translation initiation factor 3, subunit E. |
| 7 | 7684 | Hs.12013 ATP-binding cassette, sub-family E (OABP), member 1. |
| 5 | 5488 | Hs.505735 Nascent polypeptide-associated complex alpha subunit. |
| 5 | 5488 | Hs.180877 H3 histone, family 3B (H3.3B). |

EST Sequences

934 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

[Download Sequences](#)

Click the button to save sequences to a file (FASTA format)

Keywords

- blood
- normal
- unknown developmental stage
- white blood cell
- non-normalized
- bulk

Source: NCBI Clone Registry

References

[Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence](#). Adams MD, et al., *Nature* 377, 3-174 (1995).

UniGene Links





Search UniGene for

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:239 (dbEST ID)

Human White blood cells

Library Description

Organism: *Homo sapiens*
 Organ: blood
 Tissue: white blood cells

Source: Venter, JC

这里给出该文库组织样本类型的描述关键词

Gene Content Analysis

911 ESTs from this library were grouped into 707 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|------|---|
| 9 | 9879 | Hs.530096 Eukaryotic translation initiation factor 3, subunit I. |
| 9 | 9879 | Hs.524741 Actin related protein 2/3 complex, subunit 3, 21kDa. |
| 7 | 7684 | Hs.75117 Interleukin enhancer binding factor 2. |
| 7 | 7684 | Hs.643544 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide. |
| 7 | 7684 | Hs.592455 Transcribed locus, strongly similar to NP_001003447.1 rpl15 gene product [Danio rerio]. |
| 7 | 7684 | Hs.405590 Eukaryotic translation initiation factor 3, subunit E. |
| 7 | 7684 | Hs.12013 ATP-binding cassette, sub-family E (OABP), member 1. |
| 5 | 5488 | Hs.505735 Nascent polypeptide-associated complex alpha subunit. |
| 5 | 5488 | Hs.180877 H3 histone, family 3B (H3.3B). |

EST Sequences

934 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

NCBI GenBank

Click the button to save sequences to a file (FASTA format)

Keywords

- blood
- normal
- unknown developmental stage
- white blood cell
- non-normalized
- bulk

Source: NCBI Clone Registry

References

[Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence](#). Adams MD, et al., Nature 377, 3-174 (1995).

UniGene Links





NCBI



UniGene

ORGANIZED VIEW OF THE TRANSCRIPTOME

Search UniGene for

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:239 (dbEST ID)

EST Sequences

934 EST sequences from this library have been submitted to the public sequence database.

Human White blood cells

Library Description

Organism: *Homo sapiens*
Organ: blood
Tissue: white blood cells

Source: Venter, JC

文库的基因分析结果
(911条EST序列与707个UniGene
相匹配-包括假定的基因)

Gene Content Analysis

911 ESTs from this library were grouped into 707 UniGene entries (putative genes) [UniGene build #236, 09-Mar-2013]. EST counts for each entry may be used to calculate an approximate expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|------|--|
| 9 | 9879 | Hs.530096 Eukaryotic translation initiation factor 3, subunit I. |
| 9 | 9879 | Hs.524741 Actin related protein 2/3 complex, subunit 3, 21kDa. |
| 7 | 7684 | Hs.75117 Interleukin enhancer binding factor 2. |
| 7 | 7684 | Hs.643544 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide. |
| 7 | 7684 | Hs.592455 Transcribed locus, strongly similar to NP_001003447.1 rpl15 gene product [<i>Danio rerio</i>]. |
| 7 | 7684 | Hs.405590 Eukaryotic translation initiation factor 3, subunit E. |
| 7 | 7684 | Hs.12013 ATP-binding cassette, sub-family E (OABP), member 1. |
| 5 | 5488 | Hs.505735 Nascent polypeptide-associated complex alpha subunit. |
| 5 | 5488 | Hs.180877 H3 histone, family 3B (H3.3B). |

Keywords

- blood
- normal
- unknown developmental stage
- white blood cell
- non-normalized
- bulk

Source: NCBI Clone Registry

References

[Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence](#). Adams MD, et al., *Nature* 377, 3-174 (1995).

UniGene Links





Search UniGene for

NCBI > Databases > UniGene > Homo sapiens > cDNA libraries

Library:239 (dbEST ID)

Human White blood cells

Library Description

Organism: *Homo sapiens*

Organ: blood

Tissue: white blood cells

Source: Venter, JC

Gene Content Analysis

911 ESTs from this library were grouped into 70 UniGene build #236, 09-Mar-2013]. EST counts for each entry may expression level in transcripts per million (TPM).

| ESTs | TPM | UniGene Entry |
|------|------|---|
| 9 | 9879 | Hs.530096 Eukaryotic translation initiation factor 3, subunit I. |
| 9 | 9879 | Hs.524741 Actin related protein 2/3 complex, subunit 3, 21kDa. |
| 7 | 7684 | Hs.75117 Interleukin enhancer binding factor 2. |
| 7 | 7684 | Hs.643544 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide. |
| 7 | 7684 | Hs.592455 Transcribed locus, strongly similar to NP_001003447.1 rpl15 gene product [Danio rerio]. |
| 7 | 7684 | Hs.405590 Eukaryotic translation initiation factor 3, subunit E. |
| 7 | 7684 | Hs.12013 ATP-binding cassette, sub-family E (OABP), member 1. |
| 5 | 5488 | Hs.505735 Nascent polypeptide-associated complex alpha subunit. |
| 5 | 5488 | Hs.180877 H3 histone, family 3B (H3.3B). |

EST Sequences

934 EST sequences from this library have been submitted to the public sequence database as of 09-Mar-2013.

Source: NCBI GenBank

[Download Sequences](#)

[to a file \(FASTA\)](#)

文库的基因表达水平的处理
(TPM值=每百万条ESTs中某个基因的EST数量)

- normal
- unknown developmental stage
- white blood cell
- non-normalized
- bulk

Source: NCBI Clone Registry

References

[Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence.](#) Adams MD, et al., Nature 377, 3-174 (1995).

UniGene Links



UniGene – DDD - Blood

Apply knowledge of

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-------|--------------------|------|----------|---------------|-----|-----------------------|-------------|-------------|--------------|-----------|------------------|-----------------|--------------------------|
| 1 | ID | Name | ESTs | Organism | Disease state | Sex | Developmental stage | tissue type | cell line | cell type | Data type | Group by disease | Group by tissue | Group by differentiation |
| 3 | 239 | Human White blood | 911 | Homo sap | normal | | unknown | white b | | | non-n | normal | blood | differentiated cells |
| 7 | 1731 | Hembase; Erythroid | 642 | Homo sap | normal | unk | Progenitor; E | blood | Prima | Eryth | subtra | normal | blood | differentiated cells |
| 11 | 6824 | CB | 7408 | Homo sap | normal | | fetus; stem ce | cord blood | | CD34+ hem | | normal | blood | CD34+ HSCs |
| 18 | 8613 | Soares_NPBMC | 1941 | Homo sap | normal | | unknown | lymphocyte | white | norma | normal | normal | blood | differentiated cells |
| 20 | 8892 | NIH_MGC_106 | 6110 | Homo sap | normal | | unknown | natura | cell lin | white | non-n | normal | blood | differentiated cells |
| 21 | 8975 | Hembase; Erythroid | 4176 | Homo sap | normal | | Progenitor; EPO resp | Prima | Erythroid C | | | normal | blood | differentiated cells |
| 25 | 9724 | NIH_MGC_118 | 9322 | Homo sap | normal | | adult | leukocyte | | white blood | | normal | blood | differentiated cells |
| 26 | 11923 | Hembase; Erythroid | 6400 | Homo sap | normal | | Precursor erythroblas | Prima | Erythroid P | | | normal | blood | differentiated cells |
| 29 | 13833 | NIH_MGC_191 | 5890 | Homo sap | normal | | unknown | Pooled | | | perip | normal | blood | differentiated cells |
| 31 | 18364 | CD34C1 | 360 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 32 | 18366 | D9OST2 | 4236 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 33 | 18372 | CD34C2 | 533 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 34 | 18373 | CD34C3 | 536 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 35 | 18380 | CORDB2 | 533 | Homo sap | normal | | neonate | cord blood | | | normal | blood | delete | |
| 36 | 18390 | D3OST3 | 493 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 37 | 18393 | D6OST1 | 361 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 38 | 18394 | D6OST2 | 535 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 39 | 18401 | NETRP2 | 8844 | Homo sap | normal | | unknown | white b | cell lin | neutrophils | | normal | blood | differentiated cells |
| 41 | 18432 | D3OST1 | 358 | Homo sap | normal | | neonate; stem | cord bl | cell lin | CD34+ cells | | normal | blood | CD34+ HSCs |
| 42 | 18494 | PEBLM2 | 7732 | Homo sap | normal | | unknown | | prima | peripheral b | | normal | blood | differentiated cells |



UniGene – DDD - Bone marrow

Apply knowledge of

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-------|-------------|------|--------------|---------------------|--------|---------------------|--------------------------|---------------------|-----------------|------------------|------------------|-----------------|--------------------------|
| | ID | Name | ESTs | Organism | Disease state | Sex | Developmental stage | tissue type | cell line | cell type | Data type | Group by disease | Group by tissue | Group by differentiation |
| 1 | | | | | | | | | | | | | | |
| 2 | 599 | KG1-a | 653 | Homo sapiens | leukemia | male | adult | bone | KG1- | promyelocytic | non-normal | leukemia | bone | KG1-a |
| 3 | 736 | Bone marrow | 320 | Homo sapiens | normal | mixed | adult | bone marrow | | | non-normal | normal | bone | differentiated |
| 4 | 931 | Human bone | 5365 | Homo sapiens | normal | mixed | mixed | bone marrow | stromal | non-normal | normal | bone | differentiated | |
| 5 | 933 | NCI_CGAP_H | 397 | Homo sapiens | normal | | unknown | bone marrow | flow-s | non-normal | normal | bone | CD34+ HSCs | |
| 6 | 1537 | NCI_CGAP_H | 1330 | Homo sapiens | normal | | adult | bone marrow | stem | non- | normal | bone | CD34+ HSCs | |
| 7 | 1662 | NCI_CGAP_H | 539 | Homo sapiens | leukemia | | unknown | bone marrow | CD34+ | non-normal | leukemia | bone | CD34+ HSCs | |
| 8 | 1663 | NCI_CGAP_H | 506 | Homo sapiens | normal | | unknown | bone marrow | CD34+ | non-normal | normal | bone | CD34+ HSCs | |
| 9 | 5410 | Pediatric | 7346 | Homo sapiens | leukemia (juvenile) | male | pediatric | 6 years | white blood | myeloid cell | leukemia | blood | differentiated | |
| 10 | 5566 | NIH_MGC_54 | 6563 | Homo sapiens | leukemia | | unknown | bone marrow from chronic | | | leukemia | bone | differentiated | |
| 11 | 5948 | NIH_MGC_55 | 5992 | Homo sapiens | leukemia | | unknown | bone marrow from acute | | | leukemia | bone | differentiated | |
| 12 | 6975 | BM | 2449 | Homo sapiens | normal | | unknown | Bone marrow | CD34+ hematopoietic | normal | | bone | CD34+ HSCs | |
| 13 | 6976 | MDS | 3316 | Homo sapiens | normal | | unknown | Bone marrow | CD34+ hematopoietic | normal | | bone | CD34+ HSCs | |
| 14 | 7038 | Myeloma | 1795 | Homo sapiens | multiple myeloma | male | unknown | bone marrow | myeloid | non-normal | multiple myeloma | bone | differentiated | |
| 15 | 15949 | BLPD01 | 1361 | Homo sapiens | - | | unknown | bone marrow | | | | | | |
| 16 | 15950 | LKPD01 | 2681 | Homo sapiens | - | | unknown | bone marrow | | | | | | |
| 17 | 16412 | Sugano cDNA | 3618 | Homo sapiens | leukemia | female | adult | bone marrow | K562 | | leukemia | bone | K562 | |
| 18 | 17382 | Human bone | 812 | Homo sapiens | normal | | unknown | Bone marrow | enriched | normal | | bone | differentiated | |
| 19 | 18407 | ERLTTF2 | 2016 | Homo sapiens | leukemia | male | adult | erythrocyte | TF-1 | erythroleukemic | leukemia | blood | TF-1 | |
| 20 | 20861 | Novel | 1158 | Homo sapiens | leukemia (acute) | female | adult | bone | NB4 | | leukemia | bone | NB4 | |



UniGene – DDD – Normal vs Leukemia

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in different expression levels ([More about DDD](#))

这里显示出当前设定的样本组、所选定的
EST/cDNA libraries及其包含的序列数目

Species: *Homo sapiens* (human)

[Start Over](#)

Pool A: Normal DCs

10 libraries, 51968 ESTs [Edit Pool](#)

Pool B: Leukemia DCs

11 libraries, 7227 ESTs [Edit Pool](#)

[New Pool](#)

Differential Display Results

The following genes (UniGene entries) display statistically significant differences in EST counts by the Fisher Exact Test.

| A | B | | UniGene Entry |
|----------|----------|---------------------------|--|
| Normal.. | Leukem.. | | |
| 0.0224 | 0.0004 | Hs.23118 | Carbonic anhydrase I (CA1) |
| ● | ● | | |
| 0.0000 | 0.0205 | Hs.529517 | Lactotransferrin (LTF) |
| ● | ● | | |
| 0.0001 | 0.0131 | Hs.78466 | Proteasome (prosome, macropain) 26S subunit, non-ATPase, 8 (PSMD8) |
| ● | ● | | |
| 0.0130 | 0.0004 | Hs.25647 | FBJ murine osteosarcoma viral oncogene homolog (FOS) |
| ● | ● | | |



UniGene – DDD – Normal vs Leukemia

Protect yourself

Apply knowledge of

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: *Homo sapiens* (human)

[Start Over](#)

Pool A: Normal DCs

10 libraries, 51968 ESTs [Edit Pool](#)

Pool B: Leukemia DCs

11 libraries, 7227 ESTs [Edit Pool](#)

[New Pool](#)

Differential Display Results

这里就是差异表达的EST（基因）

The following genes (UniGene entries) display statistically significant differences in EST counts by the Fisher Exact Test.

| A | B | | UniGene Entry |
|----------|----------|---------------------------|--|
| Normal.. | Leukem.. | | |
| 0.0224 | 0.0004 | Hs.23118 | Carbonic anhydrase I (CA1) |
| | | | |
| 0.0000 | 0.0205 | Hs.529517 | Lactotransferrin (LTF) |
| | | | |
| 0.0001 | 0.0131 | Hs.78466 | Proteasome (prosome, macropain) 26S subunit, non-ATPase, 8 (PSMD8) |
| | | | |
| 0.0130 | 0.0004 | Hs.25647 | FBJ murine osteosarcoma viral oncogene homolog (FOS) |
| | | | |



UniGene – DDD – Analysis results

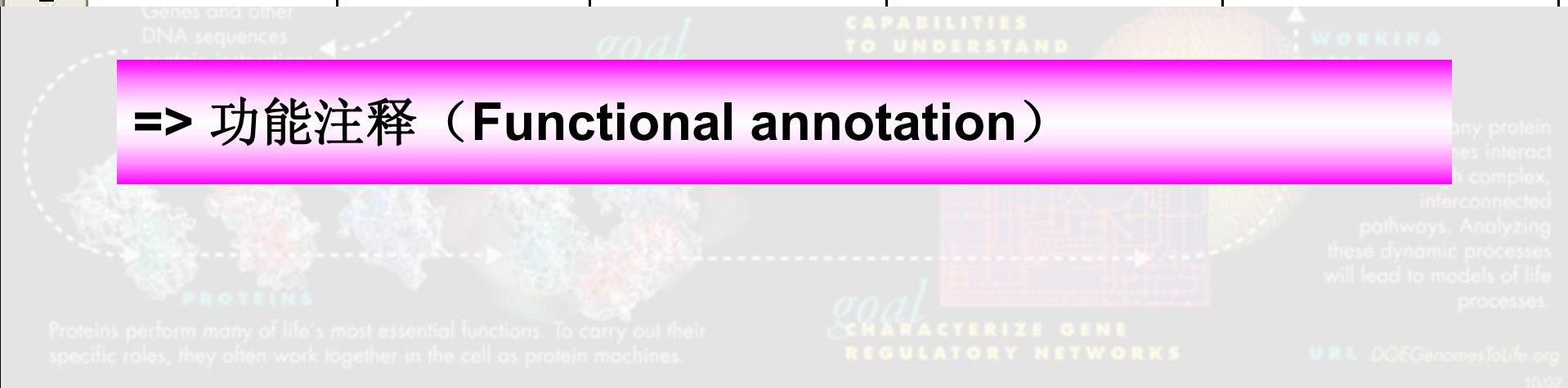
| | A | B | C | D | E |
|---|--|---|---|--|---|
| 1 | | leukemia DCs (blood) (11 libraries, 7227 ESTs) | CD34+ HSCs (9 libraries, 14820 ESTs) | Leukemia cell lines (7 libraries, 23673 ESTs) | Leukemia DCs (Bone marrow) (9 libraries, 4381 ESTs) |
| 2 | Normal DCs (blood) (10 libraries, 51968 ESTs) | 86 | 118 | 128 | 87 |



UniGene – DDD – Analysis results

| | A | B | C | D | E |
|---|---|---|---|--|---|
| 1 | | leukemia DCs (blood) (11 libraries, 7227 ESTs) | CD34+ HSCs (9 libraries, 14820 ESTs) | Leukemia cell lines (7 libraries, 23673 ESTs) | Leukemia DCs (Bone marrow) (9 libraries, 4381 ESTs) |
| 2 | | Normal DCs (blood) (10 libraries, 51968 ESTs) | 86 | 118 | 128 |

=> 功能注释 (Functional annotation)



UniGene – DDD – Analysis results

| | A | B | C | D | E |
|---|---|---|---|--|---|
| 1 | | leukemia DCs (blood) (11 libraries, 7227 ESTs) | CD34+ HSCs (9 libraries, 14820 ESTs) | Leukemia cell lines (7 libraries, 23673 ESTs) | Leukemia DCs (Bone marrow) (9 libraries, 4381 ESTs) |
| 2 | | Normal DCs (blood) (10 libraries, 51968 ESTs) | 86 | 118 | 128 |

=> 功能注释 (Functional annotation)

⇒ Search in PubMed

⇒ 是否存在有关该基因在leukemia中的研究报告



拓展阅读和实践练习

Project workers
and the public

Apply knowledge of
microbial functional
capabilities

练习EST和UniGene数据库的使用

熟悉UniGene数据库中EST文库的描述
并能够下载其序列

练习使用UniGene整合的DDD分析工具

能否使用该工具辅助探讨
生物学问题中所涉及的分子机制?

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



GENOMES to LIFE

BIOLOGICAL SOLUTIONS
FOR ENERGY CHALLENGES



如何进行基因的批量功能注释分析？

Genes and other DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY PROTEIN MACHINES

CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

WORKING CELL

Many protein machines interact through complex, interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: DOEGenomesToLife.org

10/02

Apply knowledge of microbial functional capabilities

Protect workers and the public

Clean up the environment

Sequester excess carbon

Produce and use energy



Functional Annotation

Gene Ontology (GO)

KEGG Pathway

GO and pathway annotation by DAVID

PPI network by STRING

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



GENOMES to LIFE

CELL
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES

Gene Ontology (GO)

【<http://geneontology.org/>】

Genes and other DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY PROTEIN MACHINES

CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

goal
CHARACTERIZE GENE REGULATORY NETWORKS

WORKING CELL

Many protein machines interact through complex, interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: DOEGenomesToLife.org

10/02



molecular function molecular activities of gene products

cellular component
where gene products are active

biological process
pathways and larger processes
made up of the activities of multiple
gene products.

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Apply knowledge of microbial functional capabilities

Protect workers
Clean up the environment
Sequester excess carbon
Produce and use energy

goal
EXPLORE
FUNCTIONS
IN MICROBIAL
COMMUNITIES

COMMUNITY
OF CELLS

goal
DEVELOP
COMPUTATIONAL
CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

WORKING
CELL

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: DOEGenomesToLife.org

10/02



[Contact us](#)

Enrichment analysis

Your gene IDs
here...

biological pr ▾

Homo sapien ▾

Submit

Advanced
options / Help
Powered by

Gene Ontology Consortium

Search GO data

Search

Ontology

[Filter classes](#)

[Download ontology](#)

Gene Ontology: the framework for the model of

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. [more](#)

Search documentation

Search

User stories





Contact us

Enrichment analysis

Your gene IDs
here...

biological pr ▾

Homo sapien ▾

Submit

Advanced

Gene Ontology Consortium

Search GO data

myc

Search

Ontology

Filter classes

Download ontology

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. [more](#)

Search documentation

Search





myc

Search



Text search document selection

The following results were found for **myc** using a general search over all text fields.

To narrow your search, select the type of document that you would like to search for and continue narrowing your search from the linked search page.

Ontology

Gene Ontology Term, Synonym, or Definition.

601

Genes and gene products

Genes and gene products associated with GO terms.

5331

Annotations

Associations between GO terms and genes or gene products.

6266

[Cite this data](#) • [Terms of use](#) • [GO helpdesk](#)

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

processes

URL: DOEGenomesToLife.org

10/02





Quick search

Search



Information about Ontology search

Free-text filtering



myc

Your search is pinned to these filters

+ document_category: ontology_class

No current user filters.

▶ Ontology source

▶ Subset

▶ Ancestor

▶ Obsoletion

Found entities

Total: 601; showing 1-10

Results count

10

| | Term | Definition | Ontology source | |
|--------------------------|---|---------------|-----------------|---------------------------------------|
| <input type="checkbox"/> | Mycoplasma wenyonii str. Massachusetts | | ncbi_taxonomy | Mycoplasm... |
| <input type="checkbox"/> | Mycobacterium tuberculosis H37Ra | | ncbi_taxonomy | Mycobacter... |
| <input type="checkbox"/> | Mycoplasma cynos | | ncbi_taxonomy | |
| <input type="checkbox"/> | alpha mycolate type-3 (IV) | The conjugate | chebi_ontology | (CH ₃ O) ₂ C... |



Free-text filtering

X

myc

Your search is pinned to these filters

+ document_category: ontology_class

User filters

X

- source: ncbi_taxonomy

X

- source: chebi_ontology

X

- source: none

X

- source: uberon

X

▼ Ontology source

biological_process (51) + -cell (18) + -cellular_component(11) + -molecular_function(7) + -

▶ Subset

▶ Ancestor

▶ Obsoletion

Found entities

Total: 88; showing 1-10

Results count

10

▼

1-10 <-< >> >>> ⟳ ⟲

| | Term | Definition | Ontology source | Synonyms |
|--------------------------|--------------------------------|---|--------------------|---|
| <input type="checkbox"/> | mycocerosate synthase activity | Catalysis of the reaction: acyl-CoA + 7n H(+) + methylmalonyl-CoA + 2n NADPH = n CO(2) + n CoA + n more... | molecular_function | acyl-CoA:methylmalonyl-CoA C-acetyltransferase (decarboxylating) oxoacyl-a-reducing mycocerosate synthase |
| <input type="checkbox"/> | mycodextranase activity | Catalysis of the endohydrolysis of 1,4-alpha-D-glucosidic linkages in alpha-D-glucans containing both 1,3- and 1,4-bonds. | molecular_function | 1,3-1,4-alpha-glucan 4-glucanohydrolase activity |
| <input type="checkbox"/> | mycothiol synthase activity | Catalysis of the reaction: 1-D- | molecular_function | acetyl-CoA:GlcN-Ins |

specific roles, may often work together in the cell as protein machines.

10/02



mycocerosate synthase activity

Term Information ?

Accession GO:0050111

Name mycocerosate synthase activity

Ontology molecular_function

Synonyms acyl-CoA:methylmalonyl-CoA C-acyltransferase (decarboxylating, oxoacyl- and enoyl-reducing), mycocerosic acid synthase activity

Definition Catalysis of the reaction: acyl-CoA + 7n H(+) + n methylmalonyl-CoA + 2n NADPH = n CO(2) + n CoA + n H(2)O + multi-methyl-branched acyl-CoA + 2n NADP(+). Source: EC:2.3.1.111, RHEA:10591

Comment None

History See term [history for GO:0050111](#) at QuickGO

Subset None

Community GN [Add](#) usage comments for this term on the [GONUTS](#) wiki.

Related [Link](#) to all **genes and gene products** annotated to mycocerosate synthase activity.

[Link](#) to all direct and indirect **annotations** to mycocerosate synthase activity.

[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for mycocerosate synthase activity.

Feedback Contact the [GO Helpdesk](#) if you find mistakes or have concerns about the data you find here.

Annotations

Graph Views

Inferred Tree View

Ancestors and Children

Mappings

specific roles; they often work together in the cell as protein machines.

[View All](#) [View Details](#) [View History](#) [View Log](#) [View RSS](#) [View PDF](#) [View XML](#)

10/02



Free-text filtering

X

Found entities

No results found.

Your search is pinned to these filters

- + document_category: annotation
- + regulates_closure: GO:0050111

No current user filters.

▶ Source

▶ Assigned by

▶ Ontology (aspect)

▶ Evidence type

▶ PANTHER family

▶ Qualifier

▶ Taxon

▶ Direct annotation

▶ Inferred annotation

specific roles, they often work together in the cell as protein machines.

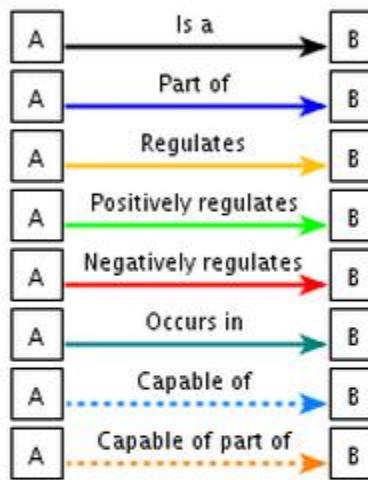
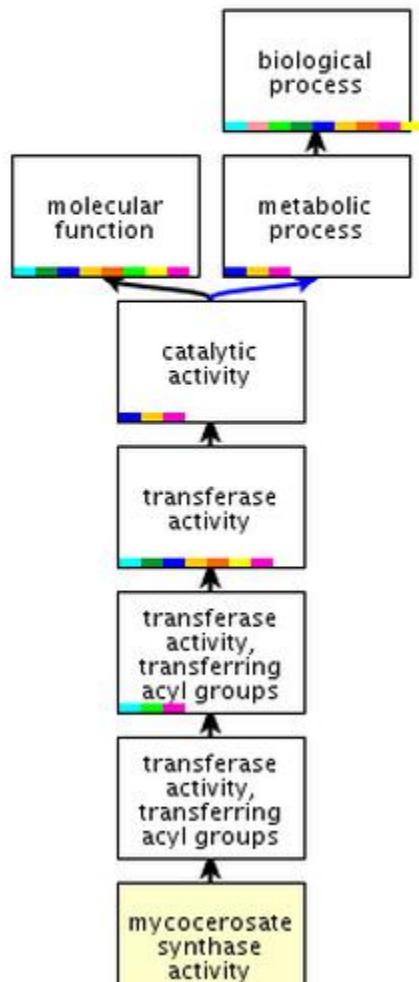
GO:0050111: regulation of transcription, promoter activity, sequence-specific DNA binding

www.genomeontology.org

10/02



[View this term in QuickGO.](#)



specific roles, may often work together in the cell as protein machines.

bioRxiv preprint doi: <https://doi.org/10.1101/2023.09.04.552912>; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

bioRxiv preprint doi: <https://doi.org/10.1101/2023.09.04.552912>; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).



- P GO:0008150 biological_process
- P GO:0008152 metabolic process
- I GO:0003674 molecular_function
- I GO:0003824 catalytic activity
- I GO:0016740 transferase activity
- I GO:0016746 transferase activity, transferring acyl groups
- I GO:0016747 transferase activity, transferring acyl groups other than amino-acyl groups
- ▼ GO:0050111 mycocerosate synthase activity

[Cite this data](#) • [Terms of use](#) • [GO helpdesk](#)

Copyright © 1999-2015 the Gene Ontology

Last file loaded on 2016-03-21, see [full details](#)

AmiGO 2 version: 2.3.2 (amigo2b)



URL: DOEGenomesToLife.org

10/02



Ancestors of mycocerosate synthase activity (GO:0050111)

subject ♦**relation** ♦**object** ♦

| | | |
|--------------------------------|--------------------|---------------------------------|
| mycocerosate synthase activity | part_of (inferred) | biological_process (GO:0008150) |
|--------------------------------|--------------------|---------------------------------|

| | | |
|--------------------------------|--------------------|--------------------------------|
| mycocerosate synthase activity | part_of (inferred) | metabolic process (GO:0008152) |
|--------------------------------|--------------------|--------------------------------|

| | | |
|--------------------------------|-----------------|---------------------------------|
| mycocerosate synthase activity | is_a (inferred) | molecular_function (GO:0003674) |
|--------------------------------|-----------------|---------------------------------|

| | | |
|--------------------------------|-----------------|---------------------------------|
| mycocerosate synthase activity | is_a (inferred) | catalytic activity (GO:0003824) |
|--------------------------------|-----------------|---------------------------------|

| | | |
|--------------------------------|-----------------|-----------------------------------|
| mycocerosate synthase activity | is_a (inferred) | transferase activity (GO:0016740) |
|--------------------------------|-----------------|-----------------------------------|

| | | |
|--------------------------------|-----------------|---|
| mycocerosate synthase activity | is_a (inferred) | transferase activity, transferring acyl groups (GO:0016746) |
|--------------------------------|-----------------|---|

| | | |
|--------------------------------|------|--|
| mycocerosate synthase activity | is_a | transferase activity, transferring acyl groups other than amino-acyl groups (GO:0016747) |
|--------------------------------|------|--|

This term has no children.



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.



will lead to models of life processes.

URL: DOEGenomesToLife.org



EC 2.3.1.111
KEGG R05189
MetaCyc MYCOCEROSATE-SYNTASE-RXN
RHEA 10591

[Cite this data](#) • [Terms of use](#) • [GO helpdesk](#)

Copyright © 1999-2015 the Gene Ontology

Last file loaded on 2016-03-21, see [full details](#)

AmiGO 2 version: 2.3.2 (amigo2b)



URL: DOEGenomesToLife.org

10/02





Contact us

Enrichment analysis

Your gene IDs
here...

biological pr ▾

Homo sapien ▾

Submit

Advanced
options / Help
Powered by

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

[Filter classes](#)

[Download ontology](#)

Gene Ontology: the framework for the model of biological systems

of the GO
is to develop

an up-to-date,
comprehensive,
**computational model of
biological systems**, from
the molecular level to
larger pathways, cellular
and organism-level
systems. [more](#)

Search documentation

Search

User stories



AmiGO 2

More information on quick search [?](#)

Quick search

Search

Get Started with Grebe



Use the Grebe Search Wizard to **get started** in exploring the Gene Ontology data.

[Go »](#)

Advanced Search



Interactively **search** the Gene Ontology data for annotations, gene products, and terms using a powerful search syntax and filters.

[Search ▾](#)

GOOSE



specific roles; they often work together in the cell as protein machines.

[View GOOSE](#) [View GOOSE](#) [View GOOSE](#) [View GOOSE](#)

[View GOOSE](#) [View GOOSE](#) [View GOOSE](#) [View GOOSE](#)

10/02



Directly query GO data using SQL

Enter your query here

Use an example query:

(Select example LEAD SQL query from the wiki)



Available mirrors:

- Berkeley BOP** : Use the LEAD full main mirror in Berkeley, CA.
- EBI** : Use the LEAD full main mirror in Cambridge, UK.
- Berkeley BOP (lite)** : Use the LEAD lite auxiliary mirror in Berkeley, CA.

Limit the number of results returned:

1000



Download results directly in a text format

Query

specific roles, they often work together in the cell as protein machines.

LEAD GOOSE GOOSE GOOSE

GO GO GO GO GO

10/02



Enrichment analysis

Tools / GO Enrichment Analysis

GO Enrichment Analysis

4072

100133623

1159

548596

1001

1750

biological pr

Homo sapien

Submit

Advanced

options / Help

Powered by
PANTHER

One of the main uses of the GO is to perform enrichment analysis on gene sets. For example, given a set of genes that are up-regulated under certain conditions, an enrichment analysis will find which GO terms are over-represented (or under-represented) using annotations for that gene set.

Enrichment analysis tool

Users can perform enrichment analyses directly from the [home page of the GOC website](#). This service connects to the analysis tool from the [PANTHER Classification System](#), which is maintained up to date with GO annotations. The PANTHER classification system is explained in great detail in [Mi H et al, PMID: 23868073](#). The list of supported gene IDs is available from the [PANTHER website](#).

Using the GO enrichment analysis tools

First, paste or type the names of the genes to be analyzed, one per row or separated by a comma. The tool can handle both MOD specific gene names and UniProt IDs (e.g. Rad54 or P38086). Second, select the name of the species (e.g. *S. cerevisiae* or *H. sapiens*) from the Species pull down and the ontology where you want to calculate the enrichment (biological process, molecular function, or cellular component). To minimize search time, the tool searches only one ontology at a time.

Interpreting the Results Table

specific roles, they often work together in the cell as protein machines.

GOC Genomes Online.org

10/02

猜你喜欢



[Home](#) [About](#) [PANTHER Data](#) [PANTHER Tools](#) [Workspace](#) [Downloads](#) [Help/Tutorial](#)

Now includes comprehensive GO annotations directly imported from the GO database

Analysis Summary: Please report in publication [?](#)

Analysis Type: PANTHER Overrepresentation Test (release 20160321)

Annotation Version and Release Date: GO Ontology database Released 2015-08-06

Analyzed List: upload_1 (Homo sapiens) [Change](#)

Reference List: Homo sapiens (all genes in database) [Change](#)

Annotation Data Set: GO biological process complete [▼](#)

Use the Bonferroni correction for multiple testing [?](#)

Results [?](#)

Reference list upload_1

Mapped IDs: [20814](#) [445](#)

Unmapped IDs: [0](#) [75](#)

specific roles, may often work together in the cell as protein machines.

[GO Consortium](#) [GO Database](#) [GO Ontologies](#) [GO Annotations](#)

10/02



Reference list upload_1

| | | |
|---------------|-----------------------|---------------------|
| Mapped IDs: | 20814 | 445 |
| Unmapped IDs: | 0 | 75 |

[Export results](#)Displaying only results with P<0.05; [click here to display all results](#)

| | Homo sapiens (REF) | # | # | expected | Fold Enrich |
|---|--------------------|-------|-----|----------|-------------|
| GO biological process complete | | | | | |
| cardiac muscle cell-cardiac muscle cell adhesion | | 7 | 5 | .15 | 33.41 |
| ↳ cell-cell adhesion | | 614 | 47 | 13.13 | 3.58 |
| ↳ cell adhesion | | 1037 | 64 | 22.17 | 2.89 |
| ↳ biological adhesion | | 1041 | 64 | 22.26 | 2.88 |
| ↳ single-organism process | | 12755 | 347 | 272.70 | 1.27 |
| regulation of cardiac muscle cell action potential | | 18 | 6 | .38 | 15.59 |
| ↳ regulation of biological process | | 10614 | 282 | 226.93 | 1.24 |
| ↳ biological regulation | | 11133 | 293 | 238.02 | 1.23 |
| homophilic cell adhesion via plasma membrane adhesion molecules | | 155 | 25 | 3.31 | 7.54 |
| ↳ cell-cell adhesion via plasma-membrane adhesion molecules | | 207 | 31 | 4.43 | 7.00 |
| endothelial cell differentiation | | 72 | 10 | 1.54 | 6.50 |

<

100

>



Explore GO data

[Home](#) allows simple quick searches using autocomplete and the new GOIr backend.

Interactively search and filter the GO data as you type. These pages allow you to use a powerful search syntax (including boolean operators).

[Annotations](#): Associations between GO terms and genes or gene products.

[Ontology](#): Gene Ontology Term, Synonym, or Definition.

[Genes and gene products](#): Genes and gene products associated with GO terms.

[Search Wizard](#) (Grebe) can be used to quickly answer common questions using a fill-in-the-blank approach.

[Visualization](#) creator for GO graph. Freely make your own ontology graph, including label and color changes. Please see the documentation for more [details](#).

Medial search returns match counts for the different search personalities and allows users to select which one they wish to continue with.

[Quick search](#)

[Search](#)

Analysis of GO data

[Term Enrichment Services](#) offers access to term enrichment analysis, finding significant shared GO terms or parents of those GO terms to help discover what input genes may have in common.

[Term Enrichment](#) [AmiGO 1.8](#) finds significant shared GO terms or parents of those GO terms, used to describe the genes in the specific roles, they often work together in the cell as protein machines.



GENOMES to LIFE

CELL
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES

Protect workers
and the public

Clean up the
environment

Apply knowledge of
microbial functional
capabilities

Produce and
use energy

KEGG

Kyoto Encyclopedia of Genes and Genomes

[\[http://www.genome.jp/kegg/\]](http://www.genome.jp/kegg/)

Genes and other
DNA sequences
contain instructions
on how and when
to build proteins

goal
IDENTIFY
PROTEIN
MACHINES

CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

WORKING
CELL

Many protein
machines interact
through complex
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life
processes.

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02





KEGG



Search

Help

» Japanese

KEGG Home[Release notes](#)
[Current statistics](#)
[Plea from KEGG](#)**KEGG Database**[KEGG overview](#)
[Searching KEGG](#)
[KEGG mapping](#)
[Color codes](#)**KEGG Objects**[Pathway maps](#)
[Brite hierarchies](#)**KEGG Software**[KegTools](#)
[KEGG API](#)
[KGML](#)**KEGG FTP**
[Subscription](#)**GenomeNet****DBGET/LinkDB**

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (See [Release notes](#) for new and updated features).

New articles

- KEGG as a reference resource for gene and protein annotation
- BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences

Main entry point to the KEGG web service**KEGG2**[KEGG Table of Contents](#)[Update notes](#)**Data-oriented entry points****KEGG PATHWAY**[KEGG pathway maps](#) [Pathway list]**KEGG BRITE**[BRITE functional hierarchies](#) [Brite list]**KEGG MODULE**[KEGG modules](#) [Module list | Statistics]**KEGG ORTHOLOGY**[Ortholog groups](#) [KO system | Annotation]**KEGG GENOME**[Genomes](#) [KEGG organisms]**KEGG GENES**[Genes and proteins](#) [Release history]**KEGG COMPOUND**[Small molecules](#) [Compound classification]

specific roles, may often work together in the cell as protein machines.

DOI:10.1101/2023.01.16.529005; version 1; available at https://doi.org/10.1101/2023.01.16.529005; this version posted January 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10/02



Database: KEGG - Search term: MYC oncogene

KEGG BRITE

ko00001
KO; KEGG Orthology (KO)
ko03000
Transcription; Transcription factors
ko03019
mRNA; Messenger RNA biogenesis

KEGG ORTHOLOGY

K04377
MYC; Myc proto-oncogene protein

K09109
NMYC, MYCN; N-myc proto-oncogene protein
K09110
LMYC, MYCL; L-myc proto-oncogene protein
K09111
BMYC, MYCB; brain expressed myelocytomatosis oncogene
K09112
SMYC, MYCS; myc-like oncogene

KEGG GENES

hsa:4609
MYC, MRTL, MYCC, bHLHe39, c-Myc; v-myc avian myelocytomatosis viral oncogene homolog; K04377 Myc proto-oncogene protein
hsa:4613
MYCN, MODED, N-myc, NMYC, ODED, bHLHe37; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog; K09109 N-myc proto-oncogene protein
hsa:4610
MYCL, L-Mvc, LMYC, MYCL1, bHLHe38; v-mvc avian myelocytomatosis viral oncogene lung carcinoma specific roles, may often work together in the cell as protein machines.



| | | |
|-------------------|--|----|
| Entry | K04377 | KO |
| Name | MYC | |
| Definition | Myc proto-oncogene protein | |
| Pathway | ko04010 MAPK signaling pathway ko04012 ErbB signaling pathway ko04110 Cell cycle ko04151 PI3K-Akt signaling pathway ko04310 Wnt signaling pathway ko04350 TGF-beta signaling pathway ko04390 Hippo signaling pathway ko04391 Hippo signaling pathway - fly ko04550 Signaling pathways regulating pluripotency of stem cells ko04630 Jak-STAT signaling pathway ko04919 Thyroid hormone signaling pathway ko05161 Hepatitis B ko05166 HTLV-I infection ko05169 Epstein-Barr virus infection ko05200 Pathways in cancer ko05202 Transcriptional misregulation in cancer ko05205 Proteoglycans in cancer ko05206 MicroRNAs in cancer ko05210 Colorectal cancer ko05213 Endometrial cancer ko05216 Thyroid cancer ko05219 Bladder cancer ko05220 Chronic myeloid leukemia ko05221 Acute myeloid leukemia ko05222 Small cell lung cancer ko05230 Central carbon metabolism in cancer | |

wledge of
functional
pathways
and
processes.

Many protein
machines interact
through complex
interconnected
ways. Analyzing
these complex
processes
can lead to models of life
processes.

GenomesToLife.org

10/02



knowledge of
biological functional
abilities

KING

Many protein machines interact through complex interconnected pathways. Analyzing dynamic processes lead to models of life processes.

DOEGenomesToLife.org

10/02

| | |
|---------|---|
| Disease | H00001 Acute lymphoblastic leukemia (ALL) (precursor B lymphoblastic leukemia) H00002 Acute lymphoblastic leukemia (ALL) (precursor T lymphoblastic leukemia) H00008 Burkitt lymphoma H00010 Multiple myeloma H00013 Small cell lung cancer H00016 Oral cancer H00025 Penile cancer H00027 Ovarian cancer H00028 Choriocarcinoma H00031 Breast cancer H00036 Osteosarcoma H00041 Kaposi's sarcoma H00055 Laryngeal cancer H01463 Mycosis fungoides |
| Brite | KEGG Orthology (KO) [BR:ko00001] Environmental Information Processing Signal transduction 04010 MAPK signaling pathway K04377 MYC; Myc proto-oncogene protein 04012 ErbB signaling pathway K04377 MYC; Myc proto-oncogene protein 04310 Wnt signaling pathway K04377 MYC; Myc proto-oncogene protein 04350 TGF-beta signaling pathway K04377 MYC; Myc proto-oncogene protein 04390 Hippo signaling pathway K04377 MYC; Myc proto-oncogene protein 04391 Hippo signaling pathway -fly K04377 MYC; Myc proto-oncogene protein 04630 Jak-STAT signaling pathway K04377 MYC; Myc proto-oncogene protein 04151 PI3K-Akt signaling pathway K04377 MYC; Myc proto-oncogene protein Cellular Processes |





Many protein machines interact through complex interconnected ways. Analyzing dynamic processes to models of life processes.

Messenger RNA biogenesis [BR:ko03019]
Eukaryotic Type
mRNA surveillance and transport factors
Factors involved in mRNA localization
Factors that have mRNA localization signals
K04377 MYC; Myc proto-oncogene protein

BRITE hierarchy

| | |
|------------------|---|
| Genes | HSA: 4609(MYC) PTR: 464393(MYC) GGO: 101137026(MYC) PON: 100456735(MYC) NLE: 100589322(MYC) MCC: 694626(MYC) MCF: 102127966(MYC) RRO: 104674798(MYC) CJC: 100407754(MYC) MMU: 17869(Myc) » show all |
| | Taxonomy KOALA UniProt |
| Reference | PMID:20399964 |
| Authors | Albihn A, Johnsen JI, Henriksson MA |
| Title | MYC in oncogenesis and as a target for cancer therapies. |
| Journal | Adv Cancer Res 107:163-224 (2010) |
| Reference | PMID:16087886 |
| Authors | Bellosta P, Hulf T, Balla Diop S, Usseglio F, Pradel J, Aragnol D, Gallant P |
| Title | Myc interacts genetically with Tip48/Reptin and Tip49/Pontin to control growth and proliferation during Drosophila development. |
| Journal | Proc Natl Acad Sci U S A 102:11799-804 (2005) |
| Sequence | [dme:Dmel_CG10798] |



| | | |
|-------------------|--|----|
| Entry | K04377 | KO |
| Name | MYC | |
| Definition | Myc proto-oncogene protein | |
| Pathway | ko04010 MAPK signaling pathway ko04012 ErbB signaling pathway ko04110 Cell cycle ko04151 PI3K-Akt signaling pathway ko04310 Wnt signaling pathway ko04350 TGF-beta signaling pathway ko04390 Hippo signaling pathway ko04391 Hippo signaling pathway - fly ko04550 Signaling pathways regulating pluripotency of stem cells ko04630 Jak-STAT signaling pathway ko04919 Thyroid hormone signaling pathway ko05161 Hepatitis B ko05166 HTLV-I infection ko05169 Epstein-Barr virus infection ko05200 Pathways in cancer ko05202 Transcriptional misregulation in cancer ko05205 Proteoglycans in cancer ko05206 MicroRNAs in cancer ko05210 Colorectal cancer ko05213 Endometrial cancer ko05216 Thyroid cancer ko05219 Bladder cancer ko05220 Chronic myeloid leukemia ko05221 Acute myeloid leukemia ko05222 Small cell lung cancer ko05230 Central carbon metabolism in cancer | |

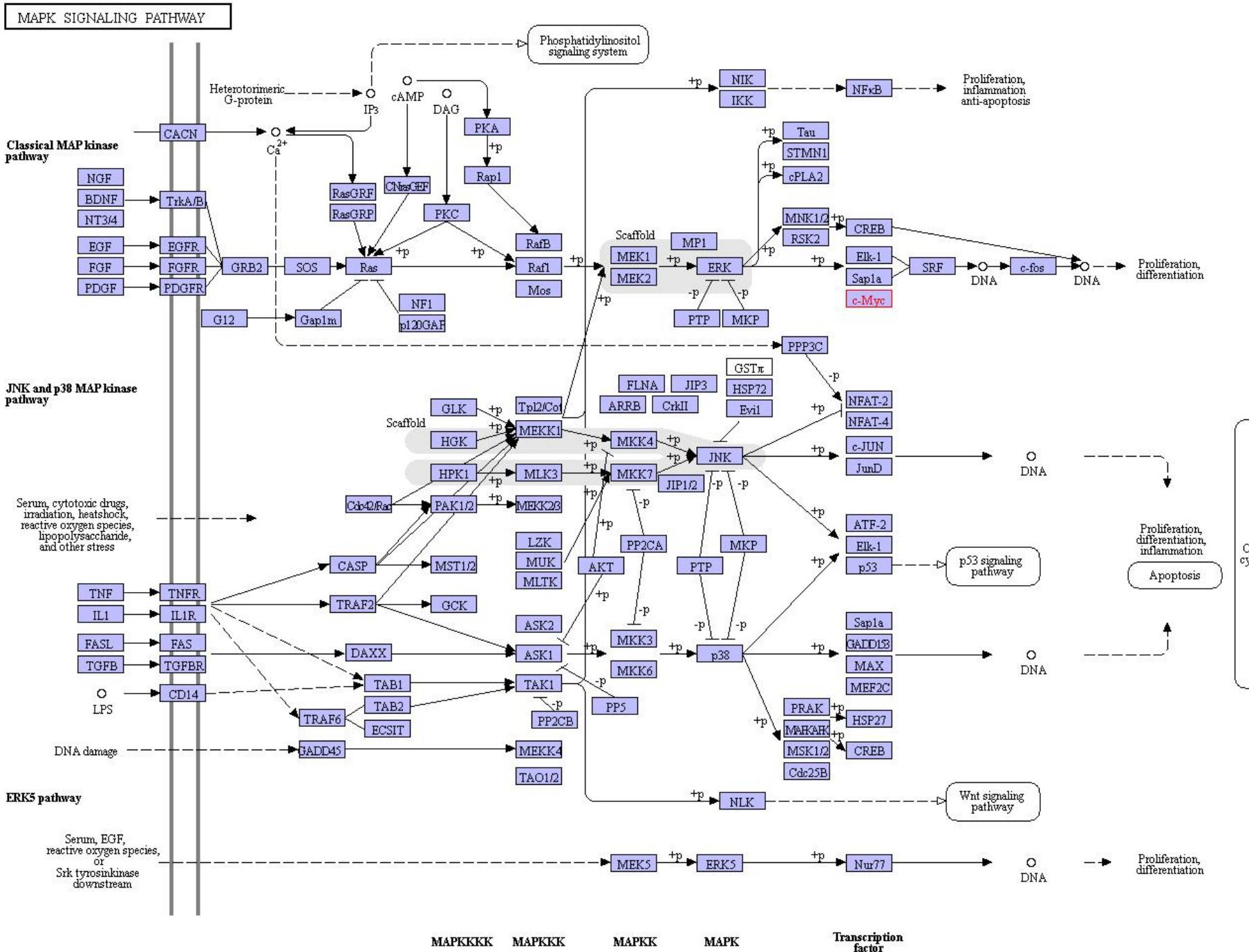
wledge of
functional
pathways
and
processes.

Many protein
machines interact
through complex
interconnected
ways. Analyzing
these complex
processes
can help us build
models of life
processes.

GenomesToLife.org

10/02





knowledge of
biological functional
abilities

KING

Many protein machines interact through complex interconnected pathways. Analyzing dynamic processes lead to models of life processes.

DOEGenomesToLife.org

10/02

| | |
|---------|---|
| Disease | H00001 Acute lymphoblastic leukemia (ALL) (precursor B lymphoblastic leukemia) H00002 Acute lymphoblastic leukemia (ALL) (precursor T lymphoblastic leukemia) H00008 Burkitt lymphoma H00010 Multiple myeloma H00013 Small cell lung cancer H00016 Oral cancer H00025 Penile cancer H00027 Ovarian cancer H00028 Choriocarcinoma H00031 Breast cancer H00036 Osteosarcoma H00041 Kaposi's sarcoma H00055 Laryngeal cancer H01463 Mycosis fungoides |
| Brite | KEGG Orthology (KO) [BR:ko00001] Environmental Information Processing Signal transduction 04010 MAPK signaling pathway K04377 MYC; Myc proto-oncogene protein 04012 ErbB signaling pathway K04377 MYC; Myc proto-oncogene protein 04310 Wnt signaling pathway K04377 MYC; Myc proto-oncogene protein 04350 TGF-beta signaling pathway K04377 MYC; Myc proto-oncogene protein 04390 Hippo signaling pathway K04377 MYC; Myc proto-oncogene protein 04391 Hippo signaling pathway -fly K04377 MYC; Myc proto-oncogene protein 04630 Jak-STAT signaling pathway K04377 MYC; Myc proto-oncogene protein 04151 PI3K-Akt signaling pathway K04377 MYC; Myc proto-oncogene protein Cellular Processes |



| | | |
|--------------------|---|--|
| Entry | H00013 | Disease |
| Name | Small cell lung cancer | |
| Description | Lung cancer is a leading cause of cancer death among men and women in industrialized countries. Small cell lung carcinoma (SCLC) is a highly aggressive neoplasm, which accounts for approximately 25% of all lung cancer cases. Molecular mechanisms altered in SCLC include induced expression of oncogene, MYC, and loss of tumorsuppressor genes, such as p53, PTEN, RB, and FHIT. The overexpression of MYC proteins in SCLC is largely a result of gene amplification. Such overexpression leads to more rapid proliferation and loss of terminal differentiation. Mutation or deletion of p53 or PTEN can lead to more rapid proliferation and reduced apoptosis. The retinoblastoma gene RB1 encodes a nuclear phosphoprotein that helps to regulate cell-cycle progression. The fragile histidine triad gene FHIT encodes the enzyme diadenosine triphosphate hydrolase, which is thought to have an indirect role in proapoptosis and cell-cycle control. | |
| Category | Cancer | |
| Brite | <p>Human diseases [BR:br08402]</p> <p>Cancers</p> <p>Cancers of the lung and pleura</p> <p>H00013 Small cell lung cancer</p> <p>Human diseases in ICD-10 classification [BR:br08403]</p> <p>2. Neoplasms (C00-D48)</p> <p>C30-C39 Malignant neoplasms of respiratory and intrathoracic organs</p> <p>C34 Malignant neoplasm of bronchus and lung</p> <p>H00013 Small cell lung cancer</p> <p>Tumor markers [br08442.html]</p> <p>Commonly used tumor markers</p> <p>H00013</p> <p>Cancer-associated carbohydrates [br08441.html]</p> <p>H00013</p> | <p>Many protein machines interact through complex interconnected ways. Analyzing dynamic processes to models of life processes.</p> <p>EGenomesToLife.org</p> <p>10/02</p> |

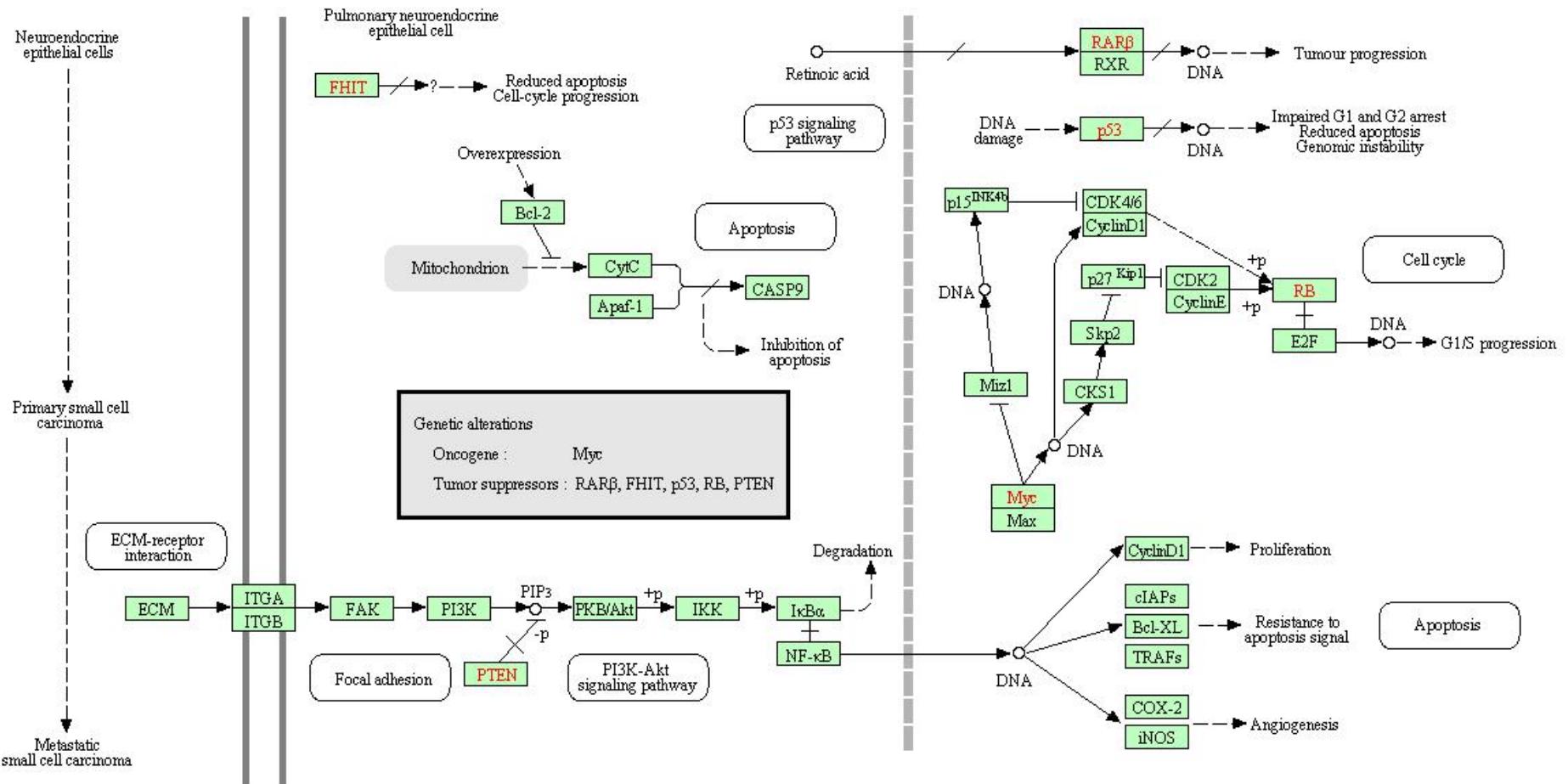


Tumor markers [br08442.html]
Commonly used tumor markers
H00013
Cancer-associated carbohydrates [br08441.html]
H00013
BRITE hierarchy

| | |
|-------------------|--|
| Pathway | hsa05222 Small cell lung cancer hsa05206 MicroRNAs in cancer |
| Gene | c-MYC (amplification) [HSA:4609] [KO:K04377] RAR-beta (promoter hypermethylation) [HSA:5915] [KO:K08528] FHIT (mutation) [HSA:2272] [KO:K01522] p53 (LOH, mutation) [HSA:7157] [KO:K04451] RB1 (LOH, mutation) [HSA:5925] [KO:K06618] PTEN (mutation, deletion) [HSA:5728] [KO:K01110] Bcl-2 (overexpression) [HSA:596] [KO:K02161] |
| Carcinogen | Arsenic and arsenic compounds [CPD:C06269] Asbestos [CPD:C16442] Beryllium and beryllium compounds [CPD:C16460] Bis(chloromethyl)ether and chloromethyl methyl ether (technical-grade) N'-Nitrosonornicotine (NNN) and 4-(N-Nitrosomethylamino)-1-(3-pyridyl)-1-butanone(NNK) [CPD:C16452 C16453] Radon-222 and its decay products Silica, crystalline (inhaled in the form of quartz or cristobalite from occupational sources) [CPD:C16459] Tobacco smoking and tobacco smoke *Aluminium production *Benzo[a]pyrene [CPD:C07535] *Cadmium and cadmium compounds *Chimney sweeping *Coal gasification *Coal-tar pitches *Coal-tars [DR:D03573] *Coke production *Furniture and cabinet making |



SMALL CELL LUNG CANCER



05222 8/7/13
(c) Kanehisa Laboratories



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

most dynamic processes
will lead to models of life
processes.

URL: DOEGenomesToLife.org



GENOMES to LIFE

CELL
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES

Protect workers
and the public

Clean up the
environment

Apply knowledge of
microbial functional
capabilities

Produce and
use energy

DAVID Functional annotation 【GO&KEGG】

<http://david.abcc.ncifcrf.gov/>

Genes and other
DNA sequences
contain instructions
on how and when
to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY
PROTEIN
MACHINES

CAPABILITIES
TO UNDERSTAND
COMPLEX
BIOLOGICAL
SYSTEMS

goal
CHARACTERIZE GENE
REGULATORY NETWORKS

WORKING
CELL

Many protein
machines interact
through complex,
interconnected
pathways. Analyzing
these dynamic processes
will lead to models of life
processes.

URL: DOEGenomesToLife.org

10/02



and the public

capabilities

Shortcut to DAVID Tools

Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)

Gene Name Batch Viewer

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedures to use DAVID!

Welcome to DAVID 6.7

2003 – 2014

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.

Produce and use energy

Search



EXPLORE IN MICROBIAL

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

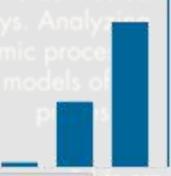
Statistics of DAVID

DAVID Bioinformatic Resources Citation

3182

Many protein through complex interconnected

pathways. Analy these dynamic proce will lead to models o



Upload List Background

Upload Gene List

[Demolist 1](#) [Demolist 2](#)

[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

Clear

Or

B: Choose From a File

Multi-List File ?

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID ▾

Step 3: List Type

Gene List

and the public

capabilities

Clean up the
environment

Functional Annotation Tool

[Tell us how you like the tool](#)
[Read technical notes of the tool](#)
[Contact us for questions](#)

BIOLOGICAL SOLUTIONS
FOR ENERGY CHALLENGES

INNOVATIVE APPROACHES
TO AN ENERGY CHALLENGE

Key Concepts: U.S. DEPARTMENT OF ENERGY

The DAVID Gene Concept

DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of millions of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)

Term/Gene Co-Occurrence Probability

Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either on the number of genes within each category or by the EASE-score. [More](#)

Gene Similarity Search

Any given gene is associating with a set of annotation terms. If genes share similar sets of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. [More](#)

Many protein
machines interact
through complex

processes



[Upload](#) [List](#) [Background](#)

Upload Gene List

[Demolist 1](#) [Demolist 2](#)
[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

```
EPCAM
CKMT1A /// CKMT1B
/// LOC100133623
CDH3
```

[Clear](#)

Or

B: Choose From a File

 [未选择文件](#)
 [Multi-List File](#)

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID ▾

Step 3: List Type

Gene List

Functional Annotation Tool

← Submit your gene list to start the tool!

[Tell us how you like the tool](#)
[Read technical notes of the tool](#)
[Contact us for questions](#)

Key Concepts:

The DAVID Gene Concept

DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of millions of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)

Term/Gene Co-Occurrence Probability

Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either on the number of genes within each category or by the EASE-score. [More](#)

Gene Similarity Search

Any given gene is associating with a set of annotation terms. If genes share similar sets of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. [More](#)



Upload Gene List

DemoList 1 DemoList 2
AFFYMETRIX_3PRIME_IVT_ID
AFFYMETRIX_EXON_GENE_ID
AFFYMETRIX_SNP_ID
AGILENT_CHIP_ID
AGILENT_ID
AGILENT_OLIGO_ID
ENSEMBL_GENE_ID
ENSEMBL_TRANSCRIPT_ID
ENTREZ_GENE_ID
FLYBASE_GENE_ID
FLYBASE_TRANSCRIPT_ID
GENBANK_ACCESSION
GENOMIC_GI_ACCESSION
GENPEPT_ACCESSION
ILLUMINA_ID
IPI_ID
MGI_ID
OFFICIAL_GENE_SYMBOL
PFAM_ID
PIR_ID
AFFYMETRIX_3PRIME_IVT_ID ▾

Step 3: List Type

Gene List
Background

Step 4: Submit List

[Submit List](#)

Functional Annotation Tool

← Submit your gene list to start the tool!

[Tell us how you like the tool](#)
[Read technical notes of the tool](#)
[Contact us for questions](#)

Key Concepts:

The DAVID Gene Concept

DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of millions of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)

Term/Gene Co-Occurrence Probability

Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either on the number of genes within each category or by the EASE-score. [More](#)

Gene Similarity Search

Any given gene is associating with a set of annotation terms. If genes share similar set of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. [More](#)

Term Similarity Search

Typically, a biological process/term is done by a corporation of a set of genes. If two or more biological processes are done by similar set of genes, the processes might be related in the biological network somehow. This search function is to identify the related biological processes/terms by quantitatively measuring the



Upload Gene List

[Demolist 1](#) [Demolist 2](#)[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

EPCAM
CKMT1A /// CKMT1B
/// LOC100133623
CDH3

[Clear](#)

Or

B: Choose From a File

[选择文件](#) [未选择文件](#) [Multi-List File](#) [?](#)

Step 2: Select Identifier

[OFFICIAL_GENE_SYMBOL](#) ▾

Step 3: List Type

 [Gene List](#) [Background](#)

Step 4: Submit List

[Submit List](#)

Functional Annotation Tool

← Submit your gene list to start the tool!

[Tell us how you like the tool](#)
[Read technical notes of the tool](#)
[Contact us for questions](#)

Key Concepts:

The DAVID Gene Concept

DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of millions of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)

Term/Gene Co-Occurrence Probability

Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either on the number of genes within each category or by the EASE-score. [More](#)

Gene Similarity Search

Any given gene is associating with a set of annotation terms. If genes share similar sets of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. [More](#)

Term Similarity Search

Typically, a biological process/term is done by a corporation of a set of genes. If two or more biological processes are done by similar sets of genes, the processes might be related in the biological network somehow. This search function is to identify the related biological processes/terms by quantitatively measuring the



Upload List Background

Gene List Manager

Select to limit annotations by one or more species

[Help](#)

- Use All Species -

Homo sapiens(485)

Mus musculus(382)

Rattus norvegicus(366)

[Select Species](#)

List Manager [Help](#)

List_1

Select List to:

[Use](#) [Rename](#)

[Remove](#) [Combine](#)

[Show Gene List](#)

[View Unmapped Ids](#)

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: List_1

3993 DAVID IDs

Current Background: Homo sapiens

Check Defaults

[Clear All](#)

- Disease (1 selected)
- Functional_Categories (3 selected)
- Gene_Ontology (3 selected)
- General Annotations (0 selected)
- Literature (0 selected)
- Main_Accessions (0 selected)
- Pathways (3 selected)
- Protein_Domains (3 selected)
- Protein_Interactions (0 selected)
- Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

[Functional Annotation Clustering](#)

[Functional Annotation Chart](#)

[Functional Annotation Table](#)



Upload List Background

Gene List Manager

Select to limit annotations by one or more species

[Help](#)

- Use All Species -

Homo sapiens(485)

Mus musculus(382)

Rattus norvegicus(366)

[Select Species](#)

[List Manager](#) [Help](#)

List_1

Select List to:

[Use](#) [Rename](#)

[Remove](#) [Combine](#)

[Show Gene List](#)

[View Unmapped Ids](#)

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: List_1

473 DAVID IDs

Current Background: Homo sapiens

Check Defaults

[Clear All](#)

- Disease (1 selected)
- Functional_Categories (3 selected)
- Gene_Ontology (3 selected)
- General Annotations (0 selected)
- Literature (0 selected)
- Main_Accessions (0 selected)
- Pathways (3 selected)
- Protein_Domains (3 selected)
- Protein_Interactions (0 selected)
- Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

[Functional Annotation Clustering](#)

[Functional Annotation Chart](#)

[Functional Annotation Table](#)



Homo sapiens(499)
Mus musculus(382)
Rattus norvegicus(366)

Select Species

List Manager Help

List_1

Select List to:

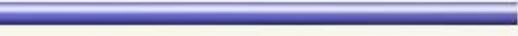
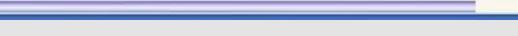
Use Rename

Remove Combine

Show Gene List

[View Unmapped Ids](#)

Gene Ontology (3 selected)

| | | | | |
|---|-------|-----|-----------------------|---|
| <input type="checkbox"/> GOTERM_BP_1 | 70.4% | 333 | Chart |  |
| <input type="checkbox"/> GOTERM_BP_2 | 70.0% | 331 | Chart |  |
| <input type="checkbox"/> GOTERM_BP_3 | 67.4% | 319 | Chart |  |
| <input type="checkbox"/> GOTERM_BP_4 | 65.1% | 308 | Chart |  |
| <input type="checkbox"/> GOTERM_BP_5 | 58.1% | 275 | Chart |  |
| <input type="checkbox"/> GOTERM_BP_ALL | 70.6% | 334 | Chart |  |
| <input checked="" type="checkbox"/> GOTERM_BP_FAT | 67.0% | 317 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_1 | 81.8% | 387 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_2 | 77.4% | 366 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_3 | 77.4% | 366 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_4 | 73.4% | 347 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_5 | 71.7% | 339 | Chart |  |
| <input type="checkbox"/> GOTERM_CC_ALL | 81.8% | 387 | Chart |  |
| <input checked="" type="checkbox"/> GOTERM_CC_FAT | 67.7% | 320 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_1 | 75.5% | 357 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_2 | 74.2% | 351 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_3 | 64.9% | 307 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_4 | 59.4% | 281 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_5 | 48.4% | 229 | Chart |  |
| <input type="checkbox"/> GOTERM_MF_ALL | 75.5% | 357 | Chart |  |
| <input checked="" type="checkbox"/> GOTERM_MF_FAT | 62.8% | 297 | Chart |  |
| <input type="checkbox"/> PANTHER_BP_ALL | 64.0% | 307 | Chart |  |





General Annotations (0 selected)

Literature (0 selected)

Main_Accessions (0 selected)

Pathways (0 selected)

Protein_Domains (0 selected)

Protein_Interactions (0 selected)

Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

[Functional Annotation Clustering](#)

[Functional Annotation Chart](#)

[Functional Annotation Table](#)



Functional Annotation Clustering

[Help and Manual](#)
Current Gene List: List_1

Current Background: Homo sapiens

473 DAVID IDs

| | | |
|-------------------------------------|---------------------------|--------------------------------|
| <input type="checkbox"/> Options | Classification Stringency | Medium ▾ |
| Rerun using options | | Create Sublist |

106 Cluster(s)
[Download File](#)

| Annotation Cluster 1 | | Enrichment Score: 5.2 | G | | Count | P_Value | Benjamini |
|--------------------------|---------------|--|----|--|-------|---------|-----------|
| <input type="checkbox"/> | GOTERM_BP_FAT | homophilic cell adhesion | RT | | 16 | 4.0E-7 | 7.8E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | cell-cell adhesion | RT | | 22 | 2.1E-6 | 2.0E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | cell adhesion | RT | | 35 | 4.3E-5 | 2.0E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | biological adhesion | RT | | 35 | 4.3E-5 | 1.7E-2 |
| Annotation Cluster 2 | | Enrichment Score: 4.79 | G | | Count | P_Value | Benjamini |
| <input type="checkbox"/> | GOTERM_CC_FAT | plasma membrane part | RT | | 96 | 1.6E-8 | 4.8E-6 |
| <input type="checkbox"/> | GOTERM_CC_FAT | intrinsic to plasma membrane | RT | | 50 | 5.0E-4 | 1.9E-2 |
| <input type="checkbox"/> | GOTERM_CC_FAT | integral to plasma membrane | RT | | 49 | 5.5E-4 | 1.8E-2 |
| Annotation Cluster 3 | | Enrichment Score: 3.81 | G | | Count | P_Value | Benjamini |
| <input type="checkbox"/> | GOTERM_CC_FAT | apical junction complex | RT | | 14 | 1.0E-6 | 1.6E-4 |
| <input type="checkbox"/> | GOTERM_CC_FAT | apicolateral plasma membrane | RT | | 14 | 1.5E-6 | 1.5E-4 |
| <input type="checkbox"/> | GOTERM_CC_FAT | cell-cell junction | RT | | 16 | 8.4E-5 | 5.1E-3 |
| <input type="checkbox"/> | GOTERM_CC_FAT | cell junction | RT | | 29 | 1.0E-4 | 5.2E-3 |
| <input type="checkbox"/> | GOTERM_CC_FAT | occluding junction | RT | | 9 | 4.5E-4 | 1.9E-2 |
| <input type="checkbox"/> | GOTERM_CC_FAT | tight junction | RT | | 9 | 4.5E-4 | 1.9E-2 |
| <input type="checkbox"/> | GOTERM_CC_FAT | desmosome | RT | | 5 | 1.3E-3 | 4.0E-2 |



Functional Annotation Chart

[Help and Manual](#)
Current Gene List: List_1

Current Background: Homo sapiens

473 DAVID IDs
 Options

[Rerun Using Options](#)
[Create Sublist](#)
[Download File](#)
139 chart records

| Sublist | Category | Term | RT | Genes | Count | % | P-Value | Benjamini |
|---------|---------------|---|----|-------|-------|------|---------|-----------|
| | GOTERM_CC_FAT | plasma membrane part | | | 96 | 20.3 | 1.6E-8 | 4.8E-6 |
| | GOTERM_BP_FAT | homophilic cell adhesion | | | 16 | 3.4 | 4.0E-7 | 7.8E-4 |
| | GOTERM_CC_FAT | apical junction complex | | | 14 | 3.0 | 1.0E-6 | 1.6E-4 |
| | GOTERM_CC_FAT | apicolateral plasma membrane | | | 14 | 3.0 | 1.5E-6 | 1.5E-4 |
| | GOTERM_BP_FAT | cell-cell adhesion | | | 22 | 4.7 | 2.1E-6 | 2.0E-3 |
| | GOTERM_CC_FAT | plasma membrane | | | 133 | 28.1 | 3.4E-6 | 2.6E-4 |
| | GOTERM_MF_FAT | calcium ion binding | | | 42 | 8.9 | 2.6E-5 | 1.4E-2 |
| | GOTERM_BP_FAT | calcium ion homeostasis | | | 16 | 3.4 | 3.5E-5 | 2.3E-2 |
| | GOTERM_BP_FAT | cell adhesion | | | 35 | 7.4 | 4.3E-5 | 2.0E-2 |
| | GOTERM_BP_FAT | biological adhesion | | | 35 | 7.4 | 4.3E-5 | 1.7E-2 |
| | GOTERM_CC_FAT | cell-cell junction | | | 16 | 3.4 | 8.4E-5 | 5.1E-3 |
| | GOTERM_BP_FAT | metal ion homeostasis | | | 16 | 3.4 | 9.6E-5 | 3.0E-2 |
| | GOTERM_CC_FAT | cell junction | | | 29 | 6.1 | 1.0E-4 | 5.2E-3 |
| | GOTERM_MF_FAT | sulfotransferase activity | | | 8 | 1.7 | 1.1E-4 | 3.0E-2 |
| | GOTERM_MF_FAT | transferase activity, transferring sulfur-containing groups | | | 8 | 1.7 | 3.3E-4 | 5.8E-2 |
| | GOTERM_BP_FAT | cellular calcium ion homeostasis | | | 14 | 3.0 | 3.7E-4 | 9.7E-2 |



Functional Annotation Table

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

3993 DAVID IDs

399 record(s)

 [Download File](#)

| DAVID ID | Gene Name | Related Genes | Homo sapiens |
|---------------|--|---------------|--------------|
| ADAMTS16 | ADAM metallopeptidase with thrombospondin type 1 motif, 16 | | |
| GOTERM_BP_FAT | proteolysis , | | |
| GOTERM_CC_FAT | extracellular region , proteinaceous extracellular matrix , extracellular matrix , extracellular region part , | | |
| GOTERM_MF_FAT | endopeptidase activity , metalloendopeptidase activity , peptidase activity , metallopeptidase activity , zinc ion binding , ion binding , cation binding , metal ion binding , transition metal ion binding , peptidase activity , acting on L-amino acid peptides , | | |
| ADAMTS19 | ADAM metallopeptidase with thrombospondin type 1 motif, 19 | | |
| GOTERM_BP_FAT | proteolysis , | | |
| GOTERM_CC_FAT | extracellular region , proteinaceous extracellular matrix , extracellular matrix , extracellular region part , | | |
| GOTERM_MF_FAT | endopeptidase activity , metalloendopeptidase activity , peptidase activity , metallopeptidase activity , zinc ion binding , ion binding , cation binding , metal ion binding , transition metal ion binding , peptidase activity , acting on L-amino acid peptides , | | |
| ARID3B | AT rich interactive domain 3B (BRIGHT-like) | | |
| GOTERM_BP_FAT | transcription , regulation of transcription , | | |
| GOTERM_MF_FAT | DNA binding , | | |
| ATP1A2 | ATPase, Na⁺/K⁺ transporting, alpha 2 (+) polypeptide | | |
| GOTERM_BP_FAT | neurotransmitter uptake , regulation of neurotransmitter levels , regulation of the force of heart contraction , regulation of respiratory gaseous exchange by neurological system process , circulatory system process , purine nucleotide metabolic process , purine nucleotide biosynthetic process , ATP biosynthetic process , ion transport , cation transport , potassium ion transport , sodium ion transport , hydrogen transport , neurotransmitter transport , cellular ion homeostasis , cellular calcium ion homeostasis , cellular metal ion homeostasis , regulation of muscle contraction , regulation of smooth muscle contraction , regulation of striated muscle contraction , cell-cell signaling , synaptic transmission , behavior , learning or memory , learning , locomotory behavior , visual behavior , blood circulation , regulation of heart contraction , regulation of blood pressure , adult locomotory behavior , visual learning , nucleoside triphosphate metabolic process , nucleoside triphosphate biosynthetic process , purine nucleoside triphosphate metabolic process , purine nucleoside triphosphate biosynthetic process , purine ribonucleotide metabolic process , purine ribonucleotide biosynthetic process , nucleotide biosynthetic process , ribonucleoside triphosphate metabolic process | | |

Gene List Manager

Select to limit annotations by one or more species

[Help](#)

- Use All Species -

Homo sapiens(485)

Mus musculus(382)

Rattus norvegicus(366)

[Select Species](#)

[List Manager](#) [Help](#)

List_1

Select List to:

[Use](#) [Rename](#)

[Remove](#) [Combine](#)

[Show Gene List](#)

[View Unmapped Ids](#)

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: List_1

473 DAVID IDs

Current Background: Homo sapiens

[Check Defaults](#)

[Clear All](#)

Disease (0 selected)

Functional_Categories (0 selected)

Gene_Ontology (0 selected)

General Annotations (0 selected)

Literature (0 selected)

Main_Accessions (0 selected)

Pathways (1 selected)

| | | | | |
|--|-------|-----|-----------------------|---|
| <input type="checkbox"/> BBID | 1.1% | 5 | Chart |  |
| <input type="checkbox"/> BIOCARTA | 4.7% | 22 | Chart |  |
| <input type="checkbox"/> EC_NUMBER | 18.2% | 86 | Chart |  |
| <input checked="" type="checkbox"/> KEGG_PATHWAY | 26.6% | 126 | Chart |  |
| <input type="checkbox"/> PANTHER_PATHWAY | 14.4% | 68 | Chart |  |
| <input type="checkbox"/> REACTOME_PATHWAY | 16.7% | 79 | Chart |  |

Protein_Domains (0 selected)

Protein_Interactions (0 selected)

Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

[Functional Annotation Clustering](#)

[Functional Annotation Chart](#)

[Functional Annotation Table](#)



Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

473 DAVID IDs

Options Classification Stringency Medium ▾

[Rerun using options](#)

[Create Sublist](#)

1 Cluster(s)

[Download File](#)

| Annotation Cluster 1 | Enrichment Score: 1.31 | G | | Count | P_Value | Benjamini |
|---------------------------------------|--|----|--|-------|---------|-----------|
| <input type="checkbox"/> KEGG_PATHWAY | Tight junction | RT | | 9 | 1.7E-2 | 8.3E-1 |
| <input type="checkbox"/> KEGG_PATHWAY | Cell adhesion molecules (CAMs) | RT | | 8 | 4.3E-2 | 9.0E-1 |
| <input type="checkbox"/> KEGG_PATHWAY | Leukocyte transendothelial migration | RT | | 6 | 1.6E-1 | 9.8E-1 |

27 terms were not clustered.

Please cite [Nature Protocols 2009; 4\(1\):44](#) & [Genome Biology 2003; 4\(5\):P3](#) within any publication that makes use of any methods inspired by DAVID.



[Term of Service](#) | [Contact Us](#) | [Site Map](#)

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

473 DAVID IDs

Options

[Rerun Using Options](#)

[Create Sublist](#)

3 chart records

[Download File](#)

| Sublist | Category | Term | RT | Genes | Count | % | P-Value | Benjamini |
|--------------------------|--------------|--|----|-------|-------|-----|---------|-----------|
| <input type="checkbox"/> | KEGG_PATHWAY | Tight junction | RT | | 9 | 1.9 | 1.7E-2 | 8.3E-1 |
| <input type="checkbox"/> | KEGG_PATHWAY | Cell adhesion molecules (CAMs) | RT | | 8 | 1.7 | 4.3E-2 | 9.0E-1 |
| <input type="checkbox"/> | KEGG_PATHWAY | Axon guidance | RT | | 7 | 1.5 | 9.7E-2 | 9.7E-1 |

454 gene(s) from your list are not in the output.

Please cite [Nature Protocols 2009; 4\(1\):44](#) & [Genome Biology 2003; 4\(5\):P3](#) within any publication that makes use of any methods inspired by DAVID.



[Term of Service](#) | [Contact Us](#) | [Site Map](#)

Functional Annotation Table

[Help and Manual](#)
Current Gene List: List_1

Current Background: Homo sapiens

3993 DAVID IDs
126 record(s)
 [Download File](#)

| ATP1A2 | ATPase, Na⁺/K⁺ transporting, alpha 2 (+) polypeptide | Related Genes | Homo sapiens |
|--------------|---|-------------------------------|------------------------------|
| KEGG_PATHWAY | Cardiac muscle contraction, Aldosterone-regulated sodium reabsorption, | Related Genes | Homo sapiens |
| BLM | Bloom syndrome, RecQ helicase-like | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Homologous recombination, | Related Genes | Homo sapiens |
| DNMT3B | DNA (cytosine-5-) methyltransferase 3 beta | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Cysteine and methionine metabolism, | Related Genes | Homo sapiens |
| DNA2 | DNA replication helicase 2 homolog (yeast) | Related Genes | Homo sapiens |
| KEGG_PATHWAY | DNA replication, | Related Genes | Homo sapiens |
| E2F5 | E2F transcription factor 5, p130-binding | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Cell cycle, TGF-beta signaling pathway, | Related Genes | Homo sapiens |
| F11R | F11 receptor | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Cell adhesion molecules (CAMs), Tight junction, Leukocyte transendothelial migration, Epithelial cell signaling in Helicobacter pylori infection, | Related Genes | Homo sapiens |
| GLI1 | GLI family zinc finger 1 | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Hedgehog signaling pathway, Pathways in cancer, Basal cell carcinoma, | Related Genes | Homo sapiens |
| KISS1R | KISS1 receptor | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Neuroactive ligand-receptor interaction, | Related Genes | Homo sapiens |
| RAB11FIP4 | RAB11 family interacting protein 4 (class II) | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Endocytosis, | Related Genes | Homo sapiens |
| RAD54L | RAD54-like (S. cerevisiae) | Related Genes | Homo sapiens |
| KEGG_PATHWAY | Homologous recombination, | Related Genes | Homo sapiens |
| RND1 | Rho family GTPase 1 | Related Genes | Homo sapiens |

GENOMES to LIFE

CELL
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES

STRING

Known and Predicted Protein-Protein Interactions

[\[http://string-db.org/\]](http://string-db.org/)

Apply knowledge of microbial functional capabilities

Protect workers and the public

Clean up the environment

Conserve energy

Produce and store energy

BIOLOGICAL
SOLUTIONS
FOR ENERGY
CHALLENGES

Genes and other DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal
IDENTIFY PROTEIN MACHINES

CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

goal
CHARACTERIZE GENE REGULATORY NETWORKS

WORKING CELL

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: DOEGenomesToLife.org

10/02



STRING - Known and Predicted Protein-Protein Interactions

search by name

protein name: (examples: #1 #2 #3)

(STRING understands a variety of protein names and accessions; you can also try a [random entry](#))

organism:
 ▼

interactors wanted:
 Proteins

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

[More Info](#)

[Funding / Support](#)

[Acknowledgements](#)

[Use Scenarios](#)

STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Franceschini et al. 2013](#) / [2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the *full* text of publications!

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)



STRING - Known and Predicted Protein-Protein Interactions

amino acid sequence: (examples: #1 #2 #3)

organism:
 auto-detect ▼

interactors wanted:
 COGs Proteins

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

More Info

Funding / Support

Acknowledgements

Use Scenarios

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Franceschini et al. 2013](#) / [2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the full text of publications!

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE ONE
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)



STRING - Known and Predicted Protein-Protein Interactions

list of names: (one per line; examples: #1 #2 #3)

... or upload a file:
 未选择文件

organism:
 auto-detect ▾

interactors wanted:
 COGs Proteins

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

More Info

[Funding / Support](#)

[Acknowledgements](#)

[Use Scenarios](#)

STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Franceschini et al. 2013](#) / [2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the *full* text of publications!

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)



STRING - Known and Predicted Protein-Protein Interactions

amino acid sequences: (examples: #1 #2)

... or upload a file:
 未选择文件

organism:
 auto-detect ▾

interactors wanted:
 COGs Proteins

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



[More Info](#)

[Funding / Support](#)

[Acknowledgements](#)

[Use Scenarios](#)

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Franceschini et al. 2013](#) / [2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the *full* text of publications!

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

Proteins perform many of the most essential functions in a cell. To carry out their specific roles, they often work together in the cell as protein machines.

REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)



STRING - Known and Predicted Protein-Protein Interactions

list of names: (one per line; examples: #1 #2 #3)

```
EPCAM
CKMT1A /// CKMT1B /// LOC100133623
CDH3
NELL2
PROM1
... or upload a file:
 未选择文件
```

organism:

Interactors wanted:

COGs Proteins

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

please enter your protein of interest...

[More Info](#)

[Funding / Support](#)

[Acknowledgements](#)

[Use Scenarios](#)

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Franceschini et al. 2013](#) / [2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the *full* text of publications!

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](#)



The following proteins in *Homo sapiens* appear to match your input.
Please review the list, then click 'Continue' to proceed.

[-> Back](#)[Continue ->](#)

'EPCAM':

- EPCAM - epithelial cell adhesion molecule; May act as a physical homophilic interaction molecule between intestinal epithelial cells (IECs) and intraepithelial lymphocytes (IELs) at the mucosal epithelium for providing immunological barrier as a first line of defense against mucosal infection (By similarity)

'CKMT1A /// CKMT1B /// LOC100133623':

-- Sorry, STRING found no proteins by this name in *Homo sapiens* --

'CDH3':

- CDH3 - cadherin 3, type 1, P-cadherin (placental); Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types

- KLK13 - kallikrein-related peptidase 13

- CDH15 - cadherin 15, type 1, M-cadherin (myotubule); Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. M-cadherin is part of the myogenic program and may provide a trigger for terminal muscle differentiation

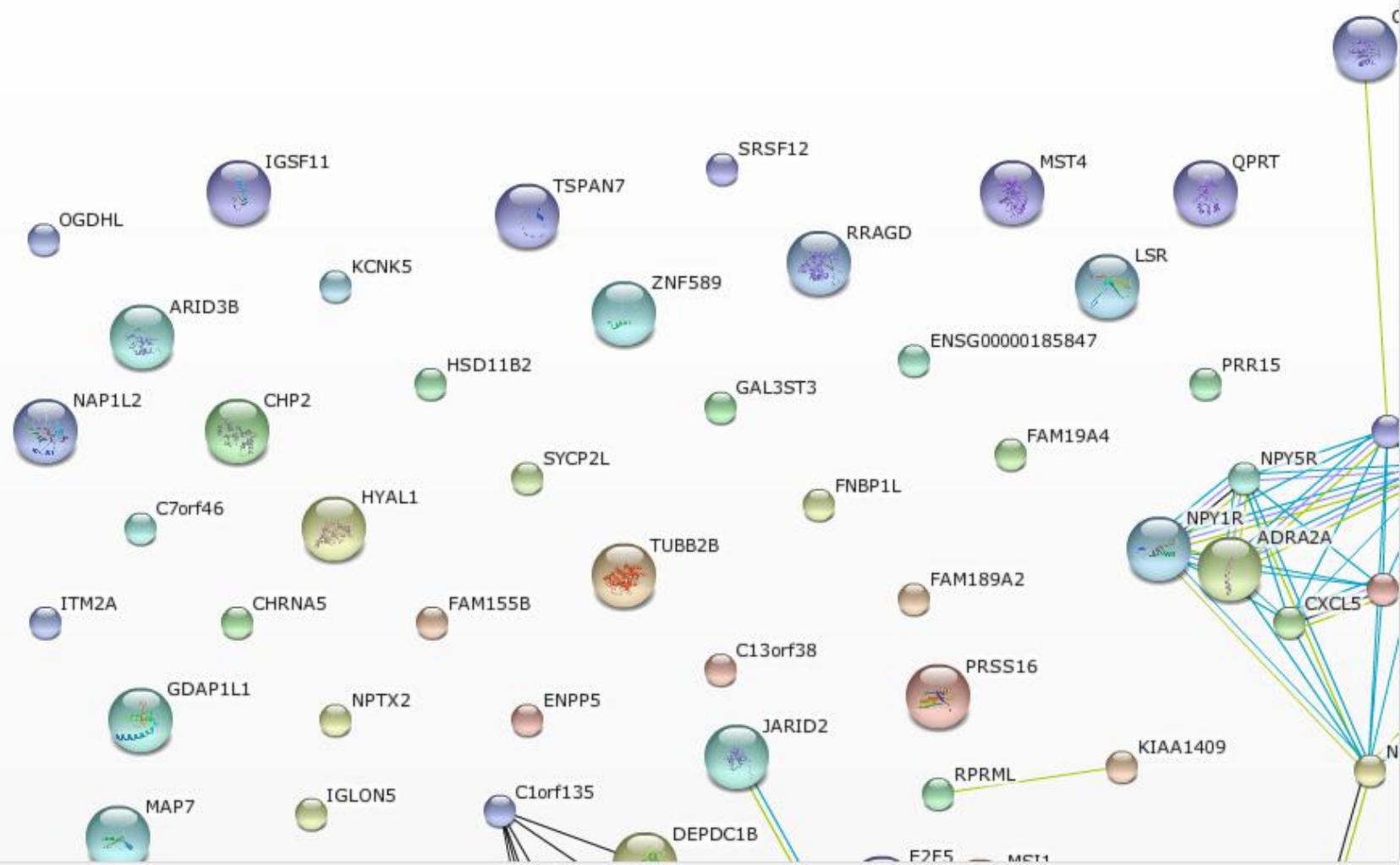
- KIAA2018 - KIAA2018

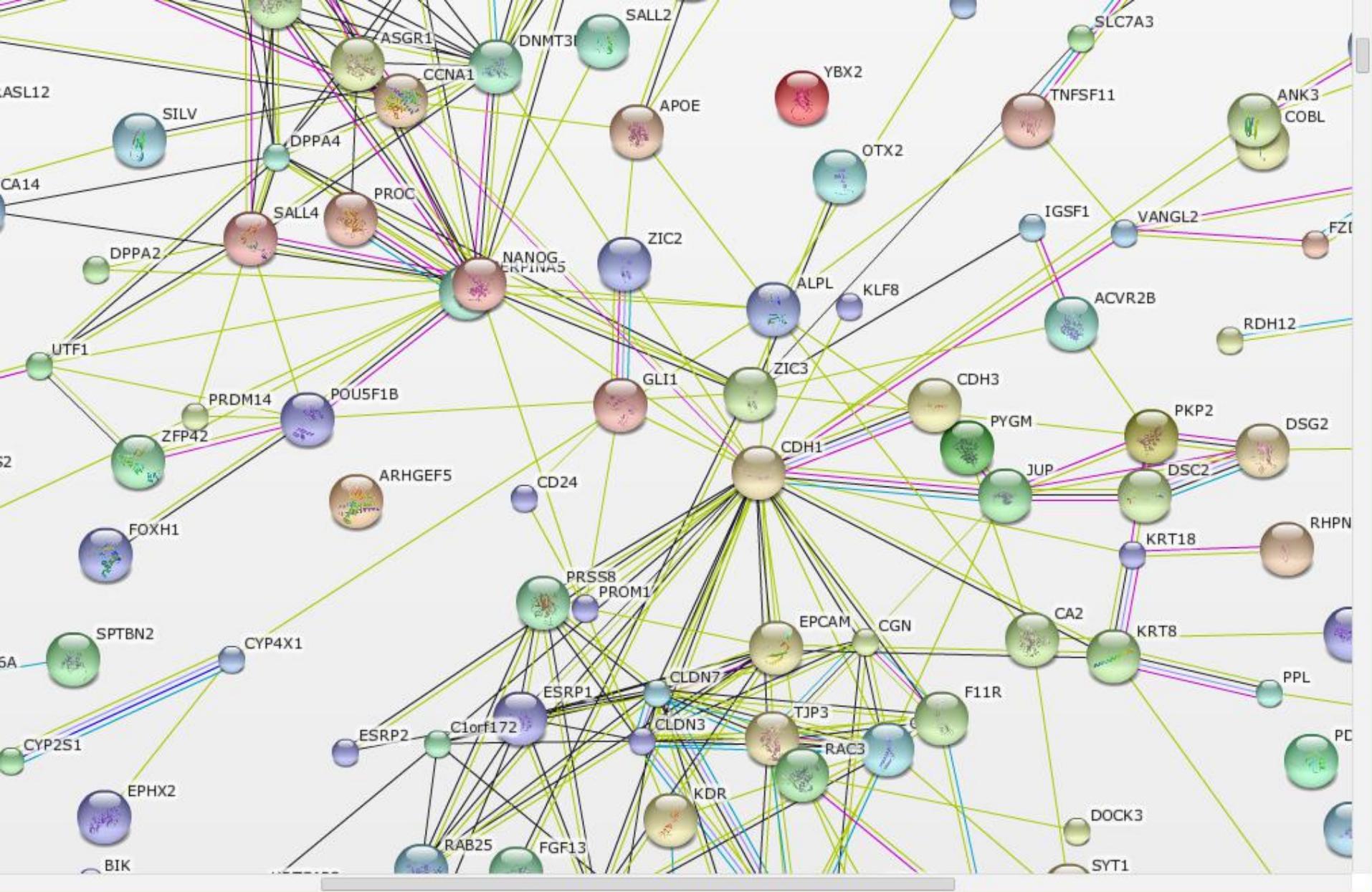
- GSN - gelsolin (amyloidosis, Finnish type); Calcium-regulated, actin-modulating protein that binds to the plus (or barbed) ends of actin monomers or filaments, preventing monomer exchange (end-blocking or capping). It can promote the assembly of monomers into filaments (nucleation) as well as sever filaments already formed

- ENSG00000235320 - microtubule-associated protein tau isoform 5

'NELL2':







拓展阅读和实践练习

Project workers
and the public

Apply knowledge of
microbial functional
capabilities

搜索并阅读更多有关**Gene ontology (GO)** 的文献资料

练习**AmiGO**等软件的使用

练习**DAVID: GO&KEGG**的使用

搜索并阅读更多有关蛋白质相互作用的文献资料

练习**STRING**的使用

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

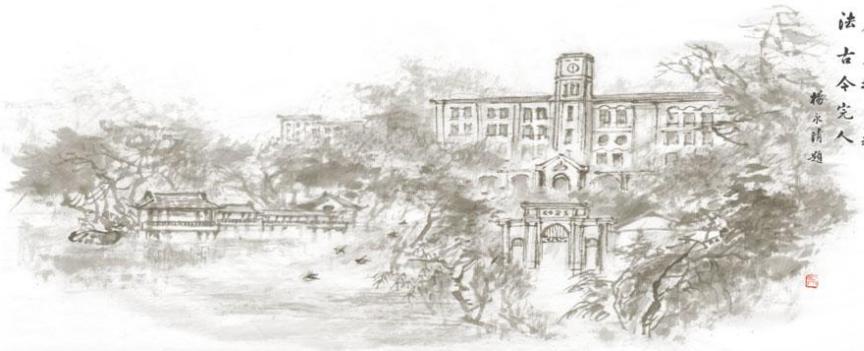
CHARACTERIZE GENE
REGULATORY NETWORKS

URL: DOEGenomesToLife.org

10/02



法養
天地正氣
古今完人
楊永清題



Thanks for your attention!

