PHỤ LỤC

TIN SINH HỌC : KHÁI NIỆM VÀ ỨNG DỤNG .

Những khái niệm cơ bản về tin sinh học (bioinformatics) .

Tin sinh học (bioinrmatics) là sự hội tụ của ba lĩnh vực công nghệ hàng đầu đó là tin học — công nghệ thông tin và công nghệ sinh học .Tin sinh học là một công cụ mới đẩy nhanh tốc độ nghiên cứu và ứng dụng của công nghệ sinh học .
Tin sinh học là khoa học bao gồm việc xây dựng , quản lý và lưu giữ nguồn dữ liệu thông tin toàn cầu , trên cơ sở đó xây dựng và hoàn thiện các chương trình xử lý dữ liệu ứng dụng làm công cụ hỗ trợ hiệu quả cho việc nghiên cứu khám phá bản chất sinh học của giới tự nhiên để phục vụ những lợi ích của con người .

Nhiệm vụ chính của tin sinh học bao gồm :

* Xây dựng , bổ sung , tổ chức và khai thác cơ sở dữ liệu đa dạng và toàn diện trên quy mô toàn cầu liên quan đến sinh học và các ngành khoa học khác .

* Xây dựng và phát triển các chương trình xử lý dữ liệu ứng dụng dưới dạng các chương trình xử lý dữ liệu độc lập được tích hợp trong các thiết bị phân tích hiện đại nhằm cung cấp cho các nhà sinh học phương tiện xây dựng phương án nghiên cứu hay phân tích xử lý kết quả với sự tham gia tư vấn của các chuyên gia trên toàn cầu .

*Đào tạo và cập nhật thường xuyên cho các nhà sinh học kỹ năng tư duy và năng lực khai thác hai nội dung trên vào hoạt động khoa học và công nghệ nhằm tạo bước chuyển biến đột phá trong cách tiếp cận và nghiên cứu thế giới sống ,tạo ra một cuộc cách mạng thực sự trong hoạt động sáng tạo của con người .

Có thể nói rằng , ngày nay các nhà khoa học ở bất kỳ một quốc gia nào cũng đều có cơ hội hoà nhập một cách bình đẳng trong nghiên cứu sinh học để mang lại những lợi ích cho cả nhân loại . Điều này có nghĩa là một nước nghèo , với trang thiết bị thông thường cũng có thể thực hiện được những chương trình nghiên cứu phức tạp thậm chí cực kỳ phức tạp nhờ sự hỗ trợ quốc tế trên mạng internet .

Cơ sở dữ liệu công nghệ sinh học .

Đặc điểm của dữ liệu công nghệ sinh học .
Nguồn cơ sở dữ liệu sinh học được truyền tải trên mạng rất đa dạng và phong phú về chủng loại cũng như khối lượng thông tin , với tốc độ ngày càng gia tăng theo thời gian . Về nội dung , cơ sở dữ liệu trải rộng trên tất cả các mặt từ các thông tin chung về tiềm lực khoa học và công nghệ của các cơ quan đến các thông tin về các công trình khoa học đã công bố , các tạp chí chuyên ngành v.v..Đặc điểm chung nhất của các dữ liệu này là được biểu diễn dưới dạng số hay ký tự trong các tệp dữ liệu đơn lẻ hay dưới dạng các chương trình thuật toán hoàn chỉnh rất thuận lợi đẻ cất giữ hay trao đổi .Nguồn tin này có thể chia thành 2 mảng lớn là dữ liệu sơ bộ và dữ liệu thứ cấp .
*Dữ liệu sơ cấp bao gồm các dữ liệu thu được qua phân tích trực tiếp bằng các phương tiện tương ứng (cơ sở dữ liệu phân tích cấu trúc DNA , cấu trúc enzym , amino axit và các chất khác)
*Dữ liệu thứ cấp gồm các dữ liệu và các thông tin thu được trên cơ sở phân tích , khái quáthoá , hệ thống hoá hay thông tin mô phỏng cho từng đối tượng hay nhóm đối tượng sinh học trong thế giới tự nhiên . Mảng dữ liệu này bao gồm cả mảng thông tin mà qua đó nhà sinh học có thể khai thác cho việc định hướng , hoạch định kế hoạch và tổ chức thực nghiệm khoa học tiếp theo sao cho hiệu quả hơn . Hoặc trên cơ sở phát triển nắm bắt được quy luật vận động của tự nhiên kết hợp với nền tảng logic của thế giơí sống có thể "thiết kế"nhữngản phẩm hoàn toàn mới , thậm chí chưa từng xuất hiện trong tự nhiên .

Một số cơ sở dữ liệu sinh học lớn trên Thế giới .

A.Dữ liệu thông tin thông thường (sách ,tạp chí ,tài liệu thông tindạng số hoá) :
-Các công trình khoa học đã công bố : PUBMED -).
-Các dữ liệu về Y —Dược (http://www.embase.com).
-Cơ sở dữ liệu nông nghiệp (http://www.nalusda.gov/general info/agricola/agricola.html) .
-Cơ sở dữ liệu về cổ sinh học và động vật hoang dã (http://www.biosis/org).
-Cơ sở dữ liệu về bệnh học trong nông nghiệp (http://www.cabi.org).

B . Dữ liệu về phân loại học (http:ww.ncbi.nlm.nih.gov/taxonomy).

C. Dữ liệu về cấu trúc và đặc tính của nucleotit và bộ gen (genome) .Có thể truy cập vào một trong 3 địa chỉ sau : http://www.ncbi.nlm.nih.gov/Genbank/index.html.
http://www.ebi,ac .uk/embl/databases.
và http://www.ddbj.nig.ac.jp.
D . Dữ liệu bộ gen người có thể truy cập vào các địa chỉ sau :OMIM:
http://www3.ncbi.nlm.nih.gov/Omim.
GDB: http://www.gdb.org.

Cơ sở dữ liệu về vi khuẩn E.Coli : http://cgsc.biology.yale.edu/top.html và http;//www.susi.bio.unigiessen.de/ecdc/ecdc.html.

Cơ sở dữ liệu về nấm men :http://www.mips.biochem.mpg.de/proj/yeast và
http://genomewww.stanford.edu/Saccharomyces.

E. Dữ liệu về cấu trúc và đặc tính chuỗi amino axit và protein :

-Protein Information Resourse PIR (http://www.nbrf.georetow.edu)
-SWISS-PROT (http:// expasy.ch) hoặc (http://www.ebi.ac.uk/swissprot).
-trEMBL (http://www.ebi.ac.uk/trEMBL)
-PROSITE (http://www.expasy.ch/prosite).
-PRINT (http://www.bioinf.man.ac.uk/bsm/dbbrowser/PRINTS/PRíNT.html).

F.Dữ liệu về proteomic (http://www.genom.ad.jp/kegg)
hoặc (http://wit.mcs.ant.gov/WIT2) hoặc (http://www..ncbi.nlm.nih.gov/COG).

G.Dữ liệu về các enzym và các con đường trao đổi chất :
ENZYM databases (http://www.expasy.ch/enzyme).
-Đặc tính enzym BRENDA (http://www..brenda.uni-koeln.de/brenda).
-Enzym và phản ứng enzym (http://www.genome.ad.jp/dbget/ligand.html).

Giới thiệu một vài trung tâm dữ liệu lớn nhất Thế giới :

TRUNG TÂM THÔNG TIN QUỐC GIA VỀ CÔNG NGHỆ SINH HỌC HOA KỲ

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

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

SITE MAP

What does NCBI do?

About NCBI

GenBank

)88 as a national resource for
y information, NCBI creates
:s, conducts research in
logy, develops software tools for
ne data, and disseminates
mation - all for the better
molecular processes affecting
disease. More...

Hot Spots

Assembly Archive

Clusters of orthologous groups

Coffee Break, Genes & Disease, NCBI Handbook

Electronic PCR

Literature databases

Molecular databases

Genomic biology

Tools

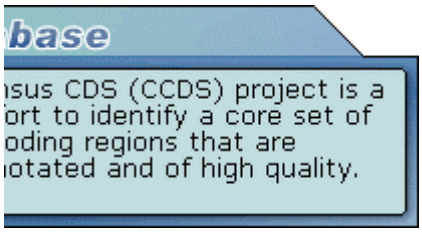
Research at NCBI

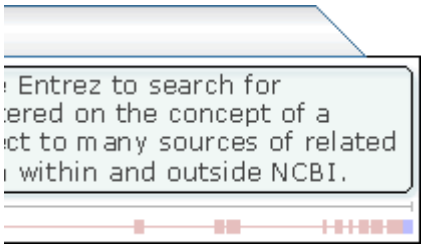
Software engineering

Education

FTP site

Contact information





News available online

ter for Biotechnology

- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Malaria genetics & genomics
- Map Viewer
- dbMHC
- Mouse genome resources
- My NCBI
- ORF finder
- Rat genome resources
- Reference sequence project
- Retrovirus resources
- SAGEmap
- SKY/CGH database
- Trace archive
- VecScreen
- NCI-CGAP

Revised: June 15, 2005.

National Center for Biotechnology Informatics (NCBI) được thành lập năm 1988 là một trong các cơ sở dữ liệu sinh học lớn nhất Thế giới . NCBI quản lý khoảng 25.10^6 nhóm dữ liệu khác nhau bao gồm các thông tin về các công trình đã công bố đến cấu trúc DNA , amino axit cũng như cấu trúc gen của các loài v.v..
Một số mảng dữ liệu lớn của trung tâm này là :
PubMed : Công bố các kết quả nghiên cứu của tất cả các tác giả . Gần đây NCBI còn có **PubMed Central** để cung cấp thêm cả những công trình khoa học đã nằm trong kế hoạch sắp phát hành để giới thiệu trước .
GenBank là mảng cơ sở dữ liệu về cấu trúc DNA và amino axit . Cơ sở dữ liệu GenBank cũng là sản phẩm quốc tế giưa 3 trung tâm dữ liệu gen lớn nhất *Thế giới là GenBank của NBCI (Hoa kỳ) , DNA Data Bank (của Nhật bản) và European Molecular Biology Laboratory nucleotide database (EMBL).*
entrez System nhằm kết nối các lien thông giữa các mảng dữ liệu giúp cho việc truy cập nhanh và đầy đủ các thông tin tìm kiếm . Tức là Entrez không phải là một cơ sở dữ liệu mà là chỉ là công cụ giúp cho người khai thác dễ dàng tiếp cận các thông tin liên quan từ nhiều mảng dữ liệu khác nhau .

CƠ SỞ DỮ LIỆU EMBL.
Phòng thí nghiệm **Sinh học phân tử Châu Âu (European Molecular Laboratory —EMBL)** được thành lập từ năm 1974 , là hệ thống liên kết các phòng thí nghiệm sinh học của 17 nước Châu Âu và Israel . Trong đó tập trung vào 5 trung tâm lớn ở **Heidelberg** và **Hambur** (Đức) , **Grenoble**(Pháp) , **Hinston** (Anh) và **Monterotondo** (Italia) . *Viện tin Sinh học Á châu (European Bioinformatics Institute , trực thuộc EMBL) được thành lập vào năm 1994 đã trở thành một trong 3 ngân hàng dữ liệu sinh học lớn nhất Thế giới .*

EMBL Outstation - The European Bioinformatics Institute

EMBL Nucleotide Sequence Database

Release Notes

Release 64 Sep 2000

EMBL Outstation
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton
Cambridge CB10 1SD

United Kingdom

Telephone: +44-1223-494400
Telefax : +44-1223-494468

Electronic mail: datalib@ebi.ac.uk
URL: http://www.ebi.ac.uk

CONTENTS

1 RELEASE 64

The EMBL Nucleotide Sequence Database was frozen to make Release 64 on
02-Sep-2000. The release contains 8,344,436 sequence entries comprising 9,650,223,037 nucleotides. This represents an increase of about 16% over Release 63. A breakdown of Release 64 by division is shown below:

Division	Entries	Nucleotides
ESTs	5,565,880	2,194,418,599
Fungi	41,017	75,333,934
GSSs	1,717,212	950,099,606
HTG	77,671	4,263,600,014
Human	119,154	965,113,287
Invertebrates	54,900	329,846,226
Other Mammals	27,021	25,376,675
Organelles	72,962	61,665,029
Patents	207,677	67,411,887
Bacteriophage	1,595	4,385,850
Plants	68,956	221,131,770
Prokaryotes	86,977	218,928,626
Rodents	55,263	92,528,729
STSs	116,671	51,039,988
Synthetic	3,838	9,763,762
Unclassified	1,174	1,869,994
Viruses	102,523	90,011,114
Other Vertebrates	23,945	27,697,947
Total	8,344,436	9,650,223,037

1.1 Nine Billion Nucleotides

On 07-JUL-2000 the number of nucleotides in the EMBL Database has passed the
9,000,000,000 mark. Over the last 12 months (compare Oct 1, 1999: 3.6 Gigabases)
the database size has increased by more than 160%.

EMBL database statistics are available at URL:
http://www3.ebi.ac.uk/Services/DBStats/

1.2 Draft Human Genome and HTG division

The completion of the human draft genome sequence has been announced on
26-June-2000. The draft sequence data is available from the EMBL Database
HTG and HUM divisions.
The total size of the euchromatic portion of the genome is estimated to be 3.2
Gbases. The fact that the total score (FIN + UNFIN) exceeds the size of the genome is due to redundancy, the general assumption is that about 30% - 40% of
the bases are redundant.
Below are the database statistics for finished and unfinished human sequence
in EMBL database from September 19, 2000.

YEAR	FIN_TOTAL	UNFIN_TOTAL	FIN + UNFIN
9/2000	910 Mb	3505 Mb	4415 Mb

See also the Genome Monitoring Table for further detailed information available from the EBI at URL
http://www.ebi.ac.uk/Databases/Genome_MOT/genome_mot.html

1.2.1 Base quality values

Quality scores from draft HTG data are available on the EBI FTP server. The gzip'ed files in the directory contain base quality values for unfinished human sequences from Japanese, US and European sequencing centres. The FastA-type headers contain the EMBL accession number/version of the corresponding database entries.

Example:

>AL009030.9 Phrap Quality (Length:229022, Min: 3, Max: 99)

In order to keep the size of the files within reasonable limits for handling purposes, files which in uncompressed form are bigger than 1 Gb, are split into smaller files.

Directory: ftp://ftp.ebi.ac.uk/pub/databases/embl/quality_scores

Current Files: /htg_sanger1.qscore.gz - /htg_sanger3.qscore.gz
/htg_genoscope1.qscore.gz
/htg_mpimg1.qscore.gz
/htg_gbf1.qscore.gz
/htg_japan1.qscore.gz
/htg_us1.qscore.gz - /htg_us9.qscore.gz

Quality score files are updated on a daily basis.

1.2.2 Ensembl automatic annotation

Ensembl provides automatic annotation to the human draft genome data including information on confirmed peptides, confirmed cDNAs and also predicted peptides.
Additionally, repeat prediction along with integration of map information and SNPs are available.
Updated human genome resources spanning the entire working draft are now available. Ensembl has released its automatic annotation for a June 15th "frozen" data set at http://freeze.ensembl.org. This URL will now be the stable location for all subsequent "frozen" dataset updates.
The Ensembl web site is available at URL http://www.ensembl.org/
Ensembl is a joint project between the Sanger Centre and EMBL-EBI.

1.3 Genome WEB Server

Access to completed genomes
The first completed genomes from viruses, phages and organelles were deposited into the EMBL Database in the early 1980's. Since then, molecular biology's shift to obtain the complete sequences of as many genomes as possible combined

with major developments in sequencing technology resulted in hundreds of complete genome sequences being added to the database, including Archaea,
Eubacteria and Eukaryota. Recent additions include Buchnera sp. APS (acc# BA000003) and Pseudomonas aeruginosa (acc# AE004091).
EBI's Genome Web Server provides easy access to completed genome sequences and
is available at URL: http://www.ebi.ac.uk/genomes/

Genome Monitoring Table
The Genome MOT presents the status of a number of large eukaryotic genome
sequencing projects. The tables are updated daily and also provide access to EMBL database entries. The Genome MOTis available at URL:
http://www.ebi.ac.uk/Databases/Genome_MOT/genome_mot.html

1.4 Cross-Reference Information

Links to a growing list of external databases have been expanded allowing integration with specialised data collections, such as protein databases, species-specific databases, taxonomy databases etc. The WWW-based sequence
retrieval system (SRS) enable users to easily navigate between cross-referenced
database entries.

EMBL links to other databases:

Database	Nr of links
RZPD	2002574
TrEMBL	338688
Demeter	175252
SWISS-PROT	143124
MaizeDB	65929
FLYBASE	40968
IMGT/LIGM	37286
MENDEL	21033
GDB	8430
MGD	7998
TRANSFAC	6620
SGD	6029
EPD	3094
IMGT/HLA	2628
Total	2859653

A list of URLs which conform with current DR line references is available:

Demeter	http://ars-genome.cornell.edu
EPD	http://www.epd.isb-sib.ch
FLYBASE	http://www.fruitfly.org
GDB	http://www.gdb.org
IMGT/HLA	http://www.ebi.ac.uk/imgt/hla
IMGT/LIGM	http://imgt.cines.fr:8104
MGD	http://www.informatics.jax.org
MaizeDB	http://www.agron.missouri.edu
MENDEL	http://mbclserver.rutgers.edu/CPGN
RZPD	http://www.rzpd.de
SGD	http://genome-www.stanford.edu
SWISS-PROT	http://www.expasy.ch
TRANSFAC	http://transfac.gbf.de/TRANSFAC
TrEMBL	http://www.ebi.ac.uk/swissprot/Information/information.html

1.5 Database Files

In order to keep the size of the data files within reasonable limits for handling purposes, additional division files will be added in subsequent releases as appropriate.

1.5.1 EST Database Files

EST files are now split according to taxonomic subdivisions following the model
of the taxonomic split of all other EMBL database divisions, e.g. Release 64
includes files

est_fun.dat Fungi ESTs
est_hum1.dat - est_hum23.dat Human ESTs
est_inv1.dat - est_inv4.dat Invertebrate ESTs
est_mam1.dat - est_mam2.dat Mammal ESTs
est_pln1.dat - est_pln8.dat Plant ESTs
est_pro.dat Prokaryote ESTs
est_rod1.dat - est_rod19.dat Rodent ESTs
est_vrt1.dat - est_vrt2.dat Vertebrate ESTs

This should reduce significantly the volume of data users have to parse in order to extract ESTs for specific groups of organisms.

1.5.2 GSS Database Files

The GSS division has been split into 18 files (gss1.dat-gss18.dat).

1.5.3 HUM Database Files

The HUM division has been split into 6 files (hum1.dat-hum6.dat).

1.5.4 HTG Database Files

The HTG division has been split into 11 files (htgo.dat and htg1.dat-htg10.dat).

htgo.dat includes all HTGS_PHASE0 entries. These typically consist of one-to-few
pass reads of a single clone, have not been assembled into contigs and are unoriented, unordered, unannotated and contain gaps with runs of 'N's separating
the reads. Low-pass sequence sampling is useful for identifying clones that may
be gene-rich. Phase0 sequences are used to check whether another center is already sequencing this clone. If not, it will be sequenced through phase 1 and
phase 2. When records are updated, the accession numbers will be preserved.
Files htg1-htg10 include all other HTG entries (HTGS_PHASE1 - HTGS_PHASE2)

1.6 Sequence Retrieval System (SRS6)

As announced earlier EBI's SRS6 server is available at URL
http://srs.ebi.ac.uk/ now maps to http://srs6.ebi.ac.uk/.

All external services are available from the Tools button on EBI's Web pages.

If you have any comments and/or suggestions please send these to:

support@ebi.ac.uk

1.7 EMBL Database FAQ

An EMBL Database FAQ has been created and is available from the EBI at URL

http://www.ebi.ac.uk/embl/Documentation/FAQ/

This document includes information on:

General questions about EMBL and other databases
Submission procedure
Updating database entries
WEBIN-specific questions
Navigation guide

1.8 Disclaimer

No guarantee is given as to the completeness and accuracy of the database entries, in particular the conformity of sequence data in the database with the journal publication where the sequence is also disclosed.

2 FORTHCOMING CHANGES

2.1 Genome Representation

At the May 2000 Collaborative Meeting it was confirmed by the sequence database collaboration DDBJ/EMBL/GenBank to go ahead to transform the currently existing experimental FTP directory representing genome data into a database division CON (Constructed Sequences) to represent complete genomes and other long sequences constructed from segment entries. The CON division entries will contain construct information (accession numbers and sequence locations) involved in building the genomes. CON entries and according information will be included into the daily data exchange mechanism between the collaborating databases.

The CON entry file includes construct information and all accession numbers relevant to the genome. Additionally, the complete entry in EMBL format (DNA and features) plus the complete DNA sequence in Fasta format is provided.
These entries will be linked, searchable and retrievable through SRS and available for BLAST and FASTA homology searching.

For an example representation, see the bacterial genome of Pseudomonas aeruginosa (AE004091) in

ftp://ftp.ebi.ac.uk/pub/databases/embl/genomes/Bacteria/paeruginosa/

AE004091.con
AE004091.embl
AE004091.embl.Z
AE004091.fasta
AE004091.fasta.Z

2.2 New HTC (High Throughput cDNA) division

At the May 2000 collaborative meeting DDBJ/EMBL/GenBank agreed to create a new database division HTC to represent unfinished High Throughput cDNA sequences. HTC sequences may include 5'UTR and 3'UTR regions and (part of a) codin region. Upon finishing of these sequences, they will be moved to the corresponding taxonomic division. HTC sequence entries will include the keyword 'HTC'. The keyword will be removed once the entry has been included in the taxonomic division.

2.3 EMBL cumulative update file

We intend to discontinue the provision of the single cumulative update file. Several sites have reported problems handling our EMBL cumulative update file when it grows beyond 2GB (uncompressed), because of file systems that do not support files > 2Gb. Instead of the cumulative.dat.gz file, we will continue to make available on our FTP server a set of smaller data files, that contain together the same data as the full cumulative update file, named cum_*.dat.gz
For further details please check the README file in directory

ftp://ftp.ebi.ac.uk/pub/databases/embl/new/

2.4 Splitting HTG and GSS division files

We plan to split HTG and GSS division files according to taxonomic subdivisions following the model of the taxonomic split of all other EMBL database divisions.
This should reduce significantly the volume of data users have to parse in order to extract HTGs and GSSs for specific groups of organisms. Files will be named accordingly e.g.
HTGS_PHASE0 sequences will be included in files htgo_hum.dat, htgo_inv.dat htgo_rod.dat etc, while htgo.dat will include all remaining HTGS_PHASE0 entries.
HTGS_PHASE1 - HTGS_PHASE2 sequences will be included in files htg_hum.dat, htg_inv.dat, htg_rod.dat etc while htg.dat will include all remaining HTG entries.
GSS sequences will be included in files gss_fun.dat, gss_hum.dat etc, while gss.dat will include all remaining GSS entries.

2.5 Next version of SRS indices

Please note that the next version of SRS indices will be for version 607x and not 606.

3 SEQUENCE SUBMISSION SYSTEMS

3.1 Checking Sequence Data For Vector Contamination

We urge submitters to remove vector contamination from sequence data before submitting to the database. To assist submitters the EBI is providing a Vector Screening Service using the latest implementation of the BLAST algorithm and a special sequence databank known as EMVEC. EMVEC is an extraction of sequences from the SYNthetic division of EMBL containing more than 2000 sequences commonly used in cloning and sequencing experiments. EMVEC is by no means a complete vector databank but EBI believes it is representative of the kind of material used in modern sequencing and should be useful to submitters. The databank will be updated with each release of EMBL and made publicly available on the EBI's ftp server for those who wish to have it.

The interactive WWW service can be found at:

http://www.ebi.ac.uk/embl/Submission/webin.html
http://www.ebi.ac.uk/blastall/vectors.html

The results will list sequences producing significant alignments and associated information like vector name, score, alignment etc

3.2 WebIn - WWW Sequence Submission System

WebIn is the preferred WWW Sequence Submission System for submitting nucleotide sequence data and associated biological information to the EMBL Nucleotide Sequence Database at the European Bioinformatics Institute(EBI). To access WebIn at the EBI please use the following URL:

http://www.ebi.ac.uk/embl/Submission/webin.html

Database entries submitted to the EMBL Nucleotide Sequence Database at the EBI will be exchanged and shared among the International Collaboration of Nucleotide Sequence Databases (DDBJ/EMBL/GenBank).

WebIn guides the user through a sequence of WWW forms allowing the submission of sequence data and descriptive information in an interactive and easy way.
All the information required to create a database entry will be collected during this process:

1 Submitter Information
2 Release Date Information
3 Sequence Data, Description and Source Information

4 Reference Citation Information
5 Feature Information (e.g. coding regions, regulators, signals etc.)

EBI staff will process data submissions within 2 working days and send the database accession number(s) assigned to your data to your e-mail address.

3.3 Bulk Submissions

With the aim to make bulk sequence submission less time consuming for the submitters, a new web-based submission system can now be accessed from the WebIn page. Authors planning to submit a large number of similar sequences (i.e.,>25) are presented with an option for "Bulk WebIn Submission". When choosing thebulk path, submitters carry on the usual WebIn submission procedure untilhaving finished a first and single representative sequence. During the submission process database staff will interactively assist in making the submission of this specific data as convenient as possible, thus saving the author the time and effort required to complete numerous submission events individually.
Alternatively, authors planning to submit very large numbers of similar sequences should contact the database before submitting the data. Database staff will create series of templates and communicate these to the author for completion with just the information unique to each sequence required.
Please contact database staff if you require further information.

e-mail: datasubs@ebi.ac.uk

Tel: +44-1223-494499
Fax: +44-1223-494472

3.4 SEQUIN - Stand-alone Submission Program

Sequin is the multi-platform (Mac/PC/Unix) stand-alone software tool developed by the NCBI for submitting entries to the EMBL, GenBank, or DDBJ sequence databases. The Sequin program, along with detailed downloading and installation instructions plus general information are available from the EBI via WWW and anonymous FTP.

http://www3.ebi.ac.uk/Services/Sequin/
ftp://ftp.ebi.ac.uk/pub/software/sequin/

3.5 Sequence Alignment Submissions

The EBI accepts submissions of alignment data (e.g. from phylogenetic and population analysis etc) of both nucleotide or amino-acid sequences, database staff assigns an alignment number (e.g. ds38200), which is then communicated to the submitter. We suggest that this number is quoted in the resulting publication.
Alignment data and associated information are made available via EBI's network

servers (see below).

ALIGNMENT FORMATS:

As well as your alignment data we require information describing your alignment
(see table below) Please provide information for all fields.

Description Field	Information required
TITLE:	Title of alignment
SUBMITTER:	Name, Affiliation, Phone, Fax, Email
RELEASE DATE:	Public Immediately / if Confidential please provide hold date
CITATION:	If known please provide complete Author list, Title, Journal, Year of publication, Page numbers
ALIGNMENT METHOD:	Method of alignment and format submitted, parameters of alignment sequences used (if appropriate)
DESCRIPTION OF SYMBOLS:	e.g. Gaps indicated by a dash '-'
DESCRIPTION OF ALIGNMENT:	Describe sequences aligned, including accession numbers (if known) and abbreviation of clones or taxon used in alignment file. If your alignment contains sequences derived from multiple taxoonomic sources, please provide the full name of each organism

FILE FORMAT:

We suggest submission in STANDARD ALIGNMENT FORMATS eg. (NEXUS, PHYLIP,
CLUSTALW etc) or Sequin output.

A sample alignment in NEXUS format can be viewed at
ftp://ftp.ebi.ac.uk/pub/databases/embl/align/ds32096.dat

NOTE 1: Alignments can be created within Sequin or imported into Sequin from
files in a standard alignment format like NEXUS or PHYLIP.

NOTE 2: If reporting new primary sequence data, we suggest that you submit
the complete individual sequence files (e.g. via Sequin or WebIn), in order to
include the sequence data as individual entries in the EMBL database. If gaps
have been introduced for the alignment, please leave them out when sending the
individual sequence files.

SENDING ALIGNMENT DATA to the EMBL Nucleotide Sequence Database
Sequence alignment data can be sent to the Nucleotide Sequence Database by
Electronic mail to datasubs@ebi.ac.uk

ACCESSING ALIGNMENT DATA
Alignment data and additional information are available via the EBI servers:

EBI WWW server:
http://www.ebi.ac.uk/embl/Submission/alignment.html
ftp://ftp.ebi.ac.uk/pub/databases/embl/align/

EBI FTP server: by anonymous FTP from FTP.EBI.AC.UK in directory pub/databases/embl/align

EBI File server: by sending an e-mail message to netserv@ebi.ac.uk including the line HELP ALIGN or GET ALIGN:DS8200.DAT

3.6 Further Submission Information

3.6.1 Annotation Guides

To help and guide submitters in annotating their sequences, two online guides are available via hyperlinks from within WebIn: EMBL Annotation Examples (http://www3.ebi.ac.uk/Services/Standards/web/) and EMBL Features and Qualifiers (http://www3.ebi.ac.uk/Services/WebFeat/). The annotation examples consist of a list of EMBL approved feature table annotations for common biological sequences. The EMBL Features and Qualifiers is a complete list of feature table key and qualifier definitions providing detailed descriptions, mandatory and optional qualifiers and usage examples.

For further information on submission of sequence data to the EMBL Nucleotide Sequence Database please access:

http://www.ebi.ac.uk/embl/Submission/

or contact database staff at:

EMBL Nucleotide Sequence Submissions
e-mail: datasubs@ebi.ac.uk
telephone: +44-1223-494499
telefax: +44-1223-494472

4 CITING THE EMBL NUCLEOTIDE SEQUENCE DATABASE

We encourage authors to include a reference to the EMBL Database in publications related to their research.

When citing data in the EMBL Database, we suggest to give the according primary accession and the publication in which the sequence first appeared.
For unpublished data, we suggest to contact the original submitters for recent publication information or revisions of the data.

We suggest to also provide a reference for the EMBL Database itself. Our recent publication describing the EMBL database should be cited:

Baker W., van den Broek, A., Camon E., Hingamp P., Sterk P., Stoesser G.,

and Tuli M.A.. 'The EMBL Nucleotide Sequence Database',
Nucl. Acids Res., 28 (1), 19-23 (2000).

Example: The numbers in parentheses refer to the REFERENCE in the EMBL
database entry, and to the EMBL citation above.

"Sequence entry X56734 (1) has been retrieved from the EMBL Database (2)
and showed significant sequence similarity to ..."

(1) Oxtoby, E., et al., Plant Mol. Biol. 17:209-219(1991).
(2) Baker, W., et al., Nucl. Acids Res. 28:19-23(2000)

5 EBI NETWORK SERVICES

5.1 Electronic Mail Server

Computer users with access to Internet (directly or via a gateway) can obtain
copies of database entries, documentation or the data submission form, by sending commands to a file server running at EBI. New and updated EMBL nucleotide sequence entries are made available on the server on a daily basis.

To use this facility, send file server commands (as electronic mail) to the address netserv@ebi.ac.uk. Each line of the mail message should consist of a
single file server request.

The most important file server request, to get started, is:

HELP

If the file server receives this command, it will return a helpfile to the sender, explaining in some detail how to use the facility. For example, to request a copy of the nucleotide sequence with accession number X55652, use
the command:

GET NUC:X55652

The file server offers various other services, (eg., access to nucleotide and protein sequence data, protein structure data, software), details of which are provided in the HELP file.

5.2 Anonymous FTP Server

An alternative method of accessing the EBI archives is to use the Internet File transfer protocol (ftp). Researchers with direct access to the Internet can use the FTP program on their local machine to connect to the host FTP.EBI.AC.UK and enter the username "anonymous" and their email address as
password.
The directory pub/help contains detailed information about the data available
from the EBI anonymous FTP server which includes the complete EMBL Nucleotide
Sequence Database releases as well as daily and weekly updates and a cumulative
update file (in UNIX-compressed format)in the following directories:

EMBL quarterly release: pub/databases/embl/release
EMBL updates: pub/databases/embl/new

5.3 World Wide Web (WWW) Server

The EBI operates a WWW server with URL http://www.ebi.ac.uk/ which gives access
to information about the EBI and it's products and services. Nucleotide sequences can be retrieved by a simple query by accession number, or more complex queries can be contructed using an SRS WWW databank browser. Nucleotide
sequences can also be submitted to the database using the interactive submission
system WebIn at URL:

http://www.ebi.ac.uk/embl/Submission/webin.html

5.4 Sequence Similarity Search Servers

The EBI offers two network servers for sequence similarity searches via electronic mail or interactive WWW forms:

FASTA based on W. Pearson's FASTA algorithm. Allows local similarity searches of protein and nucleotide sequence databases.
Send "help" to Fasta@EBI.AC.UK or use
URL http://www.ebi.ac.uk/fasta3/

BLAST based on the NCBI and WU-Blast software Send "help" to Blast@EBI.AC.UK or use URL http://www.ebi.ac.uk/blast2/

BLITZ allows very fast searches of protein sequence databases for local similarities using an exhaustive Smith-Waterman matching algorithm. Compugen's BIC_SW software is running on a Biocellerator (BIC-2) Send "help" to Blitz@EBI.AC.UK or use URL http://www.ebi.ac.uk/bic_sw/

6 DISTRIBUTION FILES

6.1 Release 64 Files

The release contains the files shown below, in the order listed. File sizes are given as numbers of records.

File Number	File Name	Description	Number of Records
1	DELETEAC.TXT	Deleted accession numbers	44649
2	FTABLE.TXT	Feature Table Documentation	465
3	RELNOTES.TXT	Release Notes (this document)	915
4	SUBFORM.TXT	Data Submission Form	418
5	SUBINFO.TXT	Data Submission Documentation	333
6	UPDATE.TXT	Data Update Form	107
7	USRMAN.TXT	User Manual	1469
8	ACNUMBER.NDX	Accession Number Index	8372365
9	CITATION.NDX	Citation Index	1872434
10	DIVISION.NDX	Division Index	23
11	KEYWORD.NDX	Keyword Index	3109242
12	SHORTDIR.NDX	Short Directory Index	21428207
13	SPECIES.NDX	Species Index	2888410
14	EST_FUN.DAT	EST Sequences	3491596
15	EST_HUM1.DAT	EST Sequences	7242162
16	EST_HUM2.DAT	EST Sequences	7383411

17	EST_HUM3.DAT	EST Sequences	7092087
18	EST_HUM4.DAT	EST Sequences	6958043
19	EST_HUM5.DAT	EST Sequences	7086795
20	EST_HUM6.DAT	EST Sequences	7098043
21	EST_HUM7.DAT	EST Sequences	7136249
22	EST_HUM8.DAT	EST Sequences	7031857
23	EST_HUM9.DAT	EST Sequences	7156374
24	EST_HUM10.DAT	EST Sequences	6859020
25	EST_HUM11.DAT	EST Sequences	6661083
26	EST_HUM12.DAT	EST Sequences	6431484
27	EST_HUM13.DAT	EST Sequences	6811351
28	EST_HUM14.DAT	EST Sequences	6856402
29	EST_HUM15.DAT	EST Sequences	7036586
30	EST_HUM16.DAT	EST Sequences	7306475
31	EST_HUM17.DAT	EST Sequences	7263236
32	EST_HUM18.DAT	EST Sequences	7357458
33	EST_HUM19.DAT	EST Sequences	7444208
34	EST_HUM20.DAT	EST Sequences	7476190
35	EST_HUM21.DAT	EST Sequences	6699624
36	EST_HUM22.DAT	EST Sequences	6963358
37	EST_HUM23.DAT	EST Sequences	4588499
38	EST_INV1.DAT	EST Sequences	6431773
39	EST_INV2.DAT	EST Sequences	6042873
40	EST_INV3.DAT	EST Sequences	6293598
41	EST_INV4.DAT	EST Sequences	4046341
42	EST_MAM1.DAT	EST Sequences	6114230
43	EST_MAM2.DAT	EST Sequences	2356039
44	EST_PLN1.DAT	EST Sequences	6750911
45	EST_PLN2.DAT	EST Sequences	6219344
46	EST_PLN3.DAT	EST Sequences	5830564
47	EST_PLN4.DAT	EST Sequences	7215994
48	EST_PLN5.DAT	EST Sequences	7046836
49	EST_PLN6.DAT	EST Sequences	6762278
50	EST_PLN7.DAT	EST Sequences	6720107
51	EST_PLN8.DAT	EST Sequences	6029205
52	EST_PRO.DAT	EST Sequences	38548
53	EST_ROD1.DAT	EST Sequences	7331559
54	EST_ROD2.DAT	EST Sequences	7567611
55	EST_ROD3.DAT	EST Sequences	7220551
56	EST_ROD4.DAT	EST Sequences	7549688
57	EST_ROD5.DAT	EST Sequences	6811012
58	EST_ROD6.DAT	EST Sequences	7086810
59	EST_ROD7.DAT	EST Sequences	9771985
60	EST_ROD8.DAT	EST Sequences	9130283
61	EST_ROD9.DAT	EST Sequences	7665029
62	EST_ROD10.DAT	EST Sequences	9177208
63	EST_ROD11.DAT	EST Sequences	9743196
64	EST_ROD12.DAT	EST Sequences	9700691
65	EST_ROD13.DAT	EST Sequences	9653685
66	EST_ROD14.DAT	EST Sequences	9473210
67	EST_ROD15.DAT	EST Sequences	9015774
68	EST_ROD16.DAT	EST Sequences	6666497
69	EST_ROD17.DAT	EST Sequences	7649778
70	EST_ROD18.DAT	EST Sequences	7420422
71	EST_ROD19.DAT	EST Sequences	738690
72	EST_VRT1.DAT	EST Sequences	7641169
73	EST_VRT2.DAT	EST Sequences	2254064
74	FUN.DAT	Fungi Sequences	3736027
75	GSS1.DAT	Genome Survey Sequences	6116578
76	GSS2.DAT	Genome Survey Sequences	6118824
77	GSS3.DAT	Genome Survey Sequences	6268149
78	GSS4.DAT	Genome Survey Sequences	6628318
79	GSS5.DAT	Genome Survey Sequences	6554451
80	GSS6.DAT	Genome Survey Sequences	6616068

81	GSS7.DAT	Genome Survey Sequences	6639716
82	GSS8.DAT	Genome Survey Sequences	6644800
83	GSS9.DAT	Genome Survey Sequences	6958158
84	GSS10.DAT	Genome Survey Sequences	6788195
85	GSS11.DAT	Genome Survey Sequences	7155659
86	GSS12.DAT	Genome Survey Sequences	6988978
87	GSS13.DAT	Genome Survey Sequences	6978243
88	GSS14.DAT	Genome Survey Sequences	6402203
89	GSS15.DAT	Genome Survey Sequences	6646868
90	GSS16.DAT	Genome Survey Sequences	7448747
91	GSS17.DAT	Genome Survey Sequences	6669805
92	GSS18.DAT	Genome Survey Sequences	1027489
93	HTG1.DAT	High Throughput Genome Sequences	7854248
94	HTG2.DAT	High Throughput Genome Sequences	5995734
95	HTG3.DAT	High Throughput Genome Sequences	4210260
96	HTG4.DAT	High Throughput Genome Sequences	4724917
97	HTG5.DAT	High Throughput Genome Sequences	8718298
98	HTG6.DAT	High Throughput Genome Sequences	8721834
99	HTG7.DAT	High Throughput Genome Sequences	8979368
100	HTG8.DAT	High Throughput Genome Sequences	8137472
101	HTG9.DAT	High Throughput Genome Sequences	7846179
102	HTG10.DAT	High Throughput Genome Sequences	4273070
103	HTGO. DAT	High Throughput Genome Sequences	8701440
104	HUM1.DAT	Human Sequences	9494007
105	HUM2.DAT	Human Sequences	5320579
106	HUM3.DAT	Human Sequences	3561983
107	HUM4.DAT	Human Sequences	2858503
108	HUM5.DAT	Human Sequences	2298449
109	HUM6.DAT	Human Sequences	1644433
110	INV.DAT	Invertebrate Sequences	9495348
111	MAM.DAT	Other Mammal Sequences	1908267
112	ORG.DAT	Organelle Sequences	5140625
113	PATENT.DAT	Patent Sequences	8110279
114	PHG.DAT	Bacteriophage Sequences	217840
115	PLN.DAT	Plant Sequences	8269953
116	PRO1.DAT	Prokaryote Sequences	6104496
117	PRO2.DAT	Prokaryote Sequences	4233076
118	ROD.DAT	Rodent Sequences	4755562
119	STS.DAT	STS Sequences	7970081
120	SYN.DAT	Synthetic Sequences	394629
121	UNC.DAT	Unclassified Sequences	106371
122	VRL.DAT	Viral Sequences	7545287
123	VRT.DAT	Other Vertebrate Sequences	1787491

6.2 SRS Indices

SRS indices can be found on the FTP server in the srs directory
ftp://ftp.ebi.ac.uk/pub/databases/embl/release/srs/.
See README file for details.
Please note that the next version of SRS indices will be for version 607x and not 606.

APPENDIX A

DATABASE GROWTH TABLE

The following table shows the growth of the EMBL Nucleotide Sequence Database at each release.

Release	Month	Entries	Nucleotides
1	06/1982	568	585433
2	04/1983	811	1114447
3	12/1983	1481	1654863
4	08/1984	1698	2147205
5	04/1985	2378	2874493
6	08/1985	4835	4567592
7	12/1985	5789	5622638
8	04/1986	6395	6353040
9	09/1986	7630	7813214
10	12/1986	8817	9766948
11	04/1987	11621	12189783
12	07/1987	12706	13638061
13	10/1987	14397	16023478
14	01/1988	15344	17272160
15	05/1988	17961	20318442
16	08/1988	19592	22625941
17	11/1988	20695	24211054
18	02/1989	22938	27249830
19	05/1989	24365	29066676
20	08/1989	26223	31240948
21	11/1989	28679	34748087
22	02/1990	31508	38165786
23	05/1990	34902	42923803
24	08/1990	37784	47354438
25	11/1990	41580	52900354
26	02/1991	43745	55859549
27	05/1991	46871	59915244
28	09/1991	54558	70448052
29	12/1991	57655	75400487
30	03/1992	63378	83574342
31	06/1992	72481	94390065
32	09/1992	79377	101292310
33	12/1992	89100	111413979
34	03/1993	99591	121420828
35	06/1993	108973	131880111
36	09/1993	127933	145401156
37	12/1993	146576	158171400
38	03/1994	167777	177550115
39	06/1994	182615	192195819
40	09/1994	209352	211017104
41	12/1994	230950	226259607
42	03/1995	303206	262559786
43	06/1995	420111	315840053
44	09/1995	506190	363273777
45	12/1995	622566	427620278
46	03/1996	701246	473691480
47	06/1996	827174	550739395
48	09/1996	928067	608931850
49	12/1996	1047263	696183789
50	03/1997	1187455	789755858
51	06/1997	1432941	931351601
52	10/1997	1787004	1181167498
53	12/1997	1917868	1281391651

54	03/1998	2125225	1427634373
55	06/1998	2330040	1607673907
56	09/1998	2689618	1904091473
57	12/1998	3046471	2164718256
58	03/1999	3272064	2355200790
59	06/1999	3952878	2924568545
60	09/1999	4719266	3543553093
61	12/1999	5303436	4508169737
62	03/2000	5865742	6120908677
63	06/2000	6760113	8255674441
64	09/2000	8344436	9650223037.

CƠ SỞ DỮ LIỆU CIB-DDBJ .
center for Information Biology and DNA Data Bank of Japan (CIB-DDBJ) là cơ sở dữ liệu của trung tâm thông tin Sinh học , viện Di truyền Quốc gia Nhật bản .

CIB —DDBJ là cơ sở dữ liệu công nghệ Sinh học quan trọng và là cơ sở dữ liệu DNA duy nhất ở Nhật bản . Bên cạnh CIB-DDBJ , viện Di truyền Quốc gia Nhật bản còn quản lý nhiều mảng dữ liệu khác cũng rất quan trọng như **World Data Center for microorganisms** (www.wdcm.nig.ac.jp) , **Genetic Resources Database** SHIGEN (www.shigen.nig.ac.jp).

CƠ SỞ DỮ LIỆU VỀ DỰ ÁN BỘ GEN NGƯỜI VÀ GEN TRỊ LIỆU .
Đây là cơ sở dữ liệu đầy đủ nhất liên quan tới Dự án bộ gen người và Gen trị liệu . Chúng ta có thể tham khảo và làm việc với bất kỳ một bộ phận nào ở bất kỳ quốc gia nào thông qua các trang Web của họ . Đây là một phương pháp tiết kiệm nhất , đồng thời cũng là hiệu quả nhất , đặc biệt với các nước mà trang thiết bị thí nghiệm còn bị hạn chế bởi sự eo hẹp về tài chính .

skip navigation

Human Genome Project Information

Gene Therapy

- Subject Index

- Send the url of this page to a friend

News
- What's New
- Meetings Calendar
- Media Guide

Basic Information
- FAQs
- Glossary
- Acronyms
- Links
- Genetics 101
- Publications

About the Project
- What is it?
- Goals
- Progress
- History
- Ethical Issues
- Benefits
- Genetics 101

Medicine & the New Genetics
- Home
- Gene Testing
- Gene Therapy
- Pharmacogenomics
- Disease Information

Quick Links to this page

What is gene therapy?

Genes, which are carried on chromosomes, are the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. Although genes get a lot of attention, it's the proteins that perform most life functions and even make up the majority of cellular structures. When genes are altered so that the encoded proteins are unable to carry out their normal functions, genetic disorders can result.

Gene therapy is a technique for correcting defective genes responsible for disease development. Researchers may use one of several approaches for correcting faulty genes:

- A normal gene may be inserted into a nonspecific location within the genome to replace a nonfunctional gene. This approach is most common.
- An abnormal gene could be swapped for a normal gene through homologous recombination.
- The abnormal gene could be repaired through selective reverse mutation, which

returns the gene to its normal function.

- The regulation (the degree to which a gene is turned on or off) of a particular gene could be altered.

How does gene therapy work?

In most gene therapy studies, a "normal" gene is inserted into the genome to replace an "abnormal," disease-causing gene. A carrier molecule called a vector must be used to deliver the therapeutic gene to the patient's target cells. Currently, the most common vector is a virus that has been genetically altered to carry normal human DNA. Viruses have evolved a way of encapsulating and delivering their genes to human cells in a pathogenic manner. Scientists have tried to take advantage of this capability and manipulate the virus genome to remove disease-causing genes and insert therapeutic genes.

Target cells such as the patient's liver or lung cells are infected with the viral vector. The vector then unloads its genetic material containing the therapeutic human gene into the target cell. The generation of a functional protein product from the therapeutic gene restores the target cell to a normal state. See a diagram depicting this process.

Some of the different types of viruses used as gene therapy vectors:

- Retroviruses - A class of viruses that can create double-stranded DNA copies of their RNA genomes. These copies of its genome can be integrated into the chromosomes of host cells. Human immunodeficiency virus (HIV) is a retrovirus.

- Adenoviruses - A class of viruses with double-stranded DNA genomes that cause respiratory, intestinal, and eye infections in humans. The virus that causes the common cold is an adenovirus.

- Adeno-associated viruses - A class of small, single-stranded DNA viruses that can insert their genetic material at a specific site on chromosome 19.

- Herpes simplex viruses - A class of double-stranded DNA viruses that infect a particular cell type, neurons. Herpes simplex virus type 1 is a common human pathogen that causes cold sores.

Sites
- Funding

Publications
- Human Genome News
- Chromosome Poster
- Primer Molecular Genetics
- To Know Ourselves
- Your Genes, Your Choices
- List of All Publications

Search This Site

search

- Contact Us
- Privacy Statement
- Site Stats and Credits

Besides virus-mediated gene-delivery systems, there are several nonviral options for gene delivery. The simplest method is the direct introduction of therapeutic DNA into target cells. This approach is limited in its application because it can be used only with certain tissues and requires large amounts of DNA.

Another nonviral approach involves the creation of an artificial lipid sphere with an aqueous core. This liposome, which carries the therapeutic DNA, is capable of passing the DNA through the target cell's membrane.

Therapeutic DNA also can get inside target cells by chemically linking the DNA to a molecule that will bind to special cell receptors. Once bound to these receptors, the therapeutic DNA constructs are engulfed by the cell membrane and passed into the interior of the target cell. This delivery system tends to be less effective than other options.

Researchers also are experimenting with introducing a 47th (artificial human) chromosome into target cells. This chromosome would exist autonomously alongside the standard 46 --not affecting their workings or causing any mutations. It would be a large vector capable of carrying substantial amounts of genetic code, and scientists anticipate that, because of its construction and autonomy, the body's immune systems would not attack it. A problem with this potential method is the difficulty in delivering such a large molecule to the nucleus of a target cell.

What is the current status of gene therapy research?

The Food and Drug Administration (FDA) has not yet approved any human gene therapy product for sale. Current gene therapy is experimental and has not proven very successful in clinical trials. Little progress has been made since the first gene therapy clinical trial began in 1990. In 1999, gene therapy suffered a major setback with the death of 18-year-old Jesse Gelsinger. Jesse was participating in a gene therapy trial for ornithine transcarboxylase deficiency (OTCD). He died from multiple organ failures 4 days after starting the treatment. His death is believed to have been triggered by a severe immune response to the adenovirus carrier.

Another major blow came in January 2003, when the FDA placed a temporary halt on all gene therapy trials using retroviral vectors in blood stem cells. FDA took this action after it learned

that a second child treated in a French gene therapy trial had developed a leukemia-like condition. Both this child and another who had developed a similar condition in August 2002 had been successfully treated by gene therapy for X-linked severe combined immunodeficiency disease (X-SCID), also known as "bubble baby syndrome."

FDA's Biological Response Modifiers Advisory Committee (BRMAC) met at the end of February 2003 to discuss possible measures that could allow a number of retroviral gene therapy trials for treatment of life-threatening diseases to proceed with appropriate safeguards. FDA has yet to make a decision based on the discussions and advice of the BRMAC meeting.

What factors have kept gene therapy from becoming an effective treatment for genetic disease?

- Short-lived nature of gene therapy - Before gene therapy can become a permanent cure for any condition, the therapeutic DNA introduced into target cells must remain functional and the cells containing the therapeutic DNA must be long-lived and stable. Problems with integrating therapeutic DNA into the genome and the rapidly dividing nature of many cells prevent gene therapy from achieving any long-term benefits. Patients will have to undergo multiple rounds of gene therapy.

- Immune response - Anytime a foreign object is introduced into human tissues, the immune system is designed to attack the invader. The risk of stimulating the immune system in a way that reduces gene therapy effectiveness is always a potential risk. Furthermore, the immune system's enhanced response to invaders it has seen before makes it difficult for gene therapy to be repeated in patients.

- Problems with viral vectors - Viruses, while the carrier of choice in most gene therapy studies, present a variety of potential problems to the patient -- toxicity, immune and inflammatory responses, and gene control and targeting issues. In addition, there is always the fear that the viral vector, once inside the patient, may recover its ability to cause disease.

- Multigene disorders - Conditions or disorders that arise from mutations in a single gene are the best candidates for

gene therapy. Unfortunately, some the most commonly occurring disorders, such as heart disease, high blood pressure, Alzheimer's disease, arthritis, and diabetes, are caused by the combined effects of variations in many genes. Multigene or multifactorial disorders such as these would be especially difficult to treat effectively using gene therapy. For more information on different types of genetic disease, see Genetic Disease Information.

What are some recent developments in gene therapy research?

- University of California, Los Angeles, research team gets genes into the brain using liposomes coated in a polymer call polyethylene glycol (PEG). The transfer of genes into the brain is a significant achievement because viral vectors are too big to get across the "blood-brain barrier." This method has potential for treating Parkinson's disease. See Undercover genes slip into the brain at NewScientist.com (March 20, 2003).

- RNA interference or gene silencing may be a new way to treat Huntington's. Short pieces of double-stranded RNA (short, interfering RNAs or siRNAs) are used by cells to degrade RNA of a particular sequence. If a siRNA is designed to match the RNA copied from a faulty gene, then the abnormal protein product of that gene will not be produced. See Gene therapy may switch off Huntington's at NewScientist.com (March 13, 2003).

- New gene therapy approach repairs errors in messenger RNA derived from defective genes. Technique has potential to treat the blood disorder thalassaemia, cystic fibrosis, and some cancers. See Subtle gene therapy tackles blood disorder at NewScientist.com (October 11, 2002).

- Gene therapy for treating children with X-SCID (sever combined immunodeficiency) or the "bubble boy" disease is stopped in France when the treatment causes leukemia in one of the patients. See 'Miracle' gene therapy trial halted at NewScientist.com (October 3, 2002).

- Researchers at Case Western Reserve University and Copernicus Therapeutics

are able to create tiny liposomes 25 nanometers across that can carry therapeutic DNA through pores in the nuclear membrane. See DNA nanoballs boost gene therapy at NewScientist.com (May 12, 2002).

- Sickle cell is successfully treated in mice. See Murine Gene Therapy Corrects Symptoms of Sickle Cell Disease from March 18, 2002, issue of The Scientist.

What are some of the ethical considerations for using gene therapy?

--Some Questions to Consider...

- What is normal and what is a disability or disorder, and who decides?
- Are disabilities diseases? Do they need to be cured or prevented?
- Does searching for a cure demean the lives of individuals presently affected by disabilities?
- Is somatic gene therapy (which is done in the adult cells of persons known to have the disease) more or less ethical than germline gene therapy (which is done in egg and sperm cells and prevents the trait from being passed on to further generations)? In cases of somatic gene therapy, the procedure may have to be repeated in future generations.
- Preliminary attempts at gene therapy are exorbitantly expensive. Who will have access to these therapies? Who will pay for their use?

Gene Therapy Links

General Information

- MEDLINEplus: Genes and Gene Therapy - Access news, information from the National Institutes of Health, clinical trials information, research, and more.
- Recombinant DNA and Gene Transfer - National Institutes of Health Guidelines
- Questions and Answers about Gene Therapy - A fact sheet from the National Cancer Institute.
- Introduction to Gene Therapy - An overview by Access Excellence.
- A Gene Therapy Primer - Introduction to

gene therapy from the bio.com.

- Gene Therapy and Your Child - From KidsHealth for Parents.
- Pioneering gene treatment gives frail toddler a new lease of life
- Gene Transfer - An overview of gene therapy science issues, ethical concerns, and regulation and policy from the Genetics & Public Policy Center.
- Cures - An introduction to gene therapy provided by discoveryhealth.com.
- Delivering the Goods - An article describing the different types of gene therapy approaches. From October 2, 2000, issue of The Scientist.
- How to Turn on a Gene - An article from Wired Magazine.
- How Viruses Are Used in Gene Therapy - From The DNA Files, a series of radio programs from SoundVision Productions.
- Human Gene Therapy: Present and Future - A Human Genome News article.
- Gene Therapy - A NewsHour with Jim Lehrer transcript covering the death of gene therapy patient, Jesse Gelsinger (February 2, 2000).
- Animations from the Tokyo Medical University Department of Paediatrics Genetics Study Group
    - Animations of Induction of Genes (Gene Therapy)
    - Animations of Problems in Gene Therapy

FDA Information

- FDA Advisory Committee Discusses Steps for Potentially Continuing Certain Gene Therapy Trials That Were Recently Placed on Hold - 2/28/2003
- FDA Places Temporary Halt On Gene Therapy Trials Using Retroviral Vectors In Blood Stem Cells - 1/14/2003
- New Initiatives to Protect Participants in Gene Therapy Trials - 3/7/2000
- Human Gene Therapy and The Role of the Food and Drug Administration - An overview from the Center for Biologics Evaluations and Research of the U.S. Food and Drug Administration.
- Human Gene Therapy Harsh Lessons, High Hopes - An article published in the September-October 2000 issue of FDA Consumer magazine.
- The Last Word: Researchers React to Gene Therapy's Pitfalls and Promises - An article published in the September-October 2000 issue of FDA Consumer magazine.
- Fundamentals of Gene Therapy - Diagrams and basic description of gene

therapy from the FDA.

Gene Therapy Ethics

- Ethical Issues in Human Gene Therapy - A Human Genome News article.
- Special Report: Ethics of Genetics - From Guardian Unlimited.
- Ethical Issues in Human Gene Therapy - A Human Genome News article.

Gene Therapy Clinical Trials

- University of Pittsburgh Molecular Medicine Institute - Contains information about ongoing and completed clinical trials.
- Gene therapy studies in ClinicalTrials.gov - The U.S. National Institutes of Health resource for public access to information on clinical research studies.
- Gene Therapy Clinical Trials - Access to a worldwide database of gene therapy clinical trials at this Web site from the publishers of The Journal of Gene Medicine. To search the database, click on "Interactive Database" at the top of the page. Access to charts, statistics, and abstracts from clinical trials results also provided.

Professional Associations

- American Society of Gene Therapy
- Australasian Gene Therapy Society (AGTS)
- European Society of Gene Therapy (ESGT)

Gene Therapy Journals (Scientific, peer-reviewed publications targeted to clinicians and researchers. Access to full-text articles in these journals typically requires a subscription.)

- Cancer Gene Therapy - From the publishers of Nature.
- Current Gene Therapy - From Bentham Science Publishers.
- Gene Therapy - From the publishers of Nature.
- Human Gene Therapy - Journal published by Mary Anne Liebert, Inc.
- The Journal of Gene Medicine - Official journal of the European Society of Gene Therapy (ESGT), Japan Society of Gene Therapy (JSGT), and the Australasian Gene Therapy Society (AGTS).
- Molecular Therapy - A monthly journal published by the American Society of

Gene Therapy (ASGT).

Other Publications

- Vector - Magazine of the Gene Therapy Center at the University of Alabama at Birmingham. Issues available for download as PDF.

Send the url of this page to a friend

To read pdf files, download the free Acrobat Reader software.

Last modified: Tuesday, October 19, 2004

Home * Contacts * Disclaimer

Base URL: www.ornl.gov/hgmis

Site sponsored by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Human Genome Program

MẪU DÒ (ĐẦU DÒ) AXIT NUCLEIC

Lai phân tử axit nucleic được dùng để chẩn đoán các trường hợp nhiễm khuẩn mà những phương pháp kinh điển không thể phát hiện được gen gây bệnh .
Để chẩn đoán sự nhiễm khuẩn và nhiều ứng dụng khác nữa , người ta đã sử dụng một thiết bị mới , đó là đầu dò axit nucleic .
Có 3 loại mẫu dò DNA : *(1) cDNA*
(2) DNA hệ gen và
(3) các nucleotit .
Cũng có thể dùng các ***mẫu dò RNA*** , nếu có sự phù hợp .
Mức độ hiệu quả phụ thuộc vào sự hiểu biết của chúng ta về trình tự gen đích . Nếu cDNA đã được xác định thì cDNA có thể sử dụng để sàng lọc thư viện gen và tách chính đoạn trình tự gen .Tức là cDNA có thể tạo ra từ quần thể và đem sử dụng , không cần tách dòng các cDNA . Kỹ thuật này thường được dùng dưới tên phương pháp *sàng lọc "cộng/trừ"* .
Nếu dòng quan tâm có chứa một đoạn chỉ biểu hiện trong những điều kiện xác định thì mẫu dò đó có thể tạo ra từ các quần thể mRNA của các tế bào biểu hiện gen này (*mẫu dò cộng*)

và từ các tế bào không biểu hiệngen này (*mẫu dò trừ*) . Bằng cách lai ghép các dòng có thể được xác định thông qua các kiểu lai của chúng với các mẫu dò cộng và trừ .
Các mẫu dò DNA hệ gen thường là các đoạn của các trình tự đã tách dòng được sử dụng như các mẫu dó dị tương đồng hoặc để xác định các dòng khác có chứa các phần bổ sung của gen nghiên cứu .
Trong kỹ thuật này Chromosome walking (NST đi bộ) và Chromosome jumping (NST nhảy) giúp cho việc xác định các đoạn trình tự gối lên nhau và sau đó nếu nối các mẫu lại sẽ có các mạch DNA dài được xác định .
Có thể dùng các ***mẫu dò oligonucleotit*** khi biết trình tự của một số axit amin trong phân tử protein được mã bởi gen đích . Dùng mã di truyền để xác định trình tự gen tương ứng và tạo ra oligonucleotit.
Ưu điểm nổi bật của mẫu dò nucleotit là chỉ cần một đoạn trình tự ngắn là mẫu dò đã rất hữu hiệu vì vậy các gen chưa tách dòng vẫn có thể xác định trình tự thông qua trình tự của các đoạn peptit và có thể thiết kế được các mẫu dò tương ứng .
Tuy nhiên do bản chất thoái hoá của mã di truyền nên ta không thể tiên đoán một cách hoàn toàn chính xác trình tự gen. Nhưng đó không phải là khó khăn lớn vì ta có thể sử dụng các mẫu dò hỗn hợp bao gồm tất cả các trình tự có thể có .
Mẫu dò axit nucleic đang trở thành một công cụ hữu hiệu trong nhiều lĩnh vực nghiên cứu và ứng dụng công nghệ sinh học . Một ứng dụng thiết thực là cho phép chẩn đoán các trường hợp nhiễm khuẩn mà không phát hiện được kháng nguyên . Ví dụ , đối với các axit nucleic của gen gây bệnh nằm ngoài NST hoặc loại hoà nhập , không được biểu hiện rõ qua kiểu hình thì phương pháp lai phân tử vẫn cho phép chẩn đoán được .
Mẫu dò đã và đang là công cụ đắc lực trong công nghệ sinh học nói chung và trong y sinh học nói riêng .

NHỮNG TÀI LIỆU THAM KHẢO CHÍNH

1. Nguyễn Văn Cách . Tin-Sinh học . Nxb . KHKT, HN ,2005.

2. Lê Đình Lương . Nguyên lý kỹ thuật Di truyền . Nxb . KHKT, HN , 2001.

3. Phan Cự Nhân .Di truyền học Động vật .Nxb. KHKT, HN, 2001.

4. Khuất Hữu Thanh . Liệu pháp gen . Nxb. KHKT , HN , 2005 .

5. Bruce A . Chabner (Editor) ; Dan L . Longo (Editor) . Cancer Chemotherapy and Biotherapy : Principles and Practice .
Lippincott Williams Wilkins Publisher , 1996.

6. I.Edward Alcamo . DNA Technology
the awasome skill . Wm. C. Brown Publisher.
Dubuque , IA Bogota Boston Buenos Aires Caracas Chicago Guilford , CT London Madrid Mexico City Sedney Toronto , 1996.

7. Jackque B. Wallach . Interpretation of Diagnostic Test 7th Edition .
Lippincott Williams & Wilkin Publisher , 2000.

8. J. Caroline . Suicide gene therapy . Springer , 2004.

*David C. Dale (Editor) . Infectious Diseases : The Clinican's Guide to Diagnosis , Treatment and Prevention . Webmd Scientific American Medicine by WebMD Professional Publishing , 2003.

9. K Fran , Md . Austen (Editor) ; michael M., Md . frank (Editor) ; Havey I., Md .Canter (Editor) ; John P.,Md . Atkinson (Editor) ; Max Samter (Editor) . Samter's Immunologic Diseases 6th Edition .
Lippincott Williams & Wilkins Publishers , 2001.

10. Robert K . Murray ; Daryl K. Granner ; Peter A. Mayer ; Victor W. Rodwell . Harper's Illustracted Biochemistry .
Lange Medical Book / Mc Graw —Hill .
Medical Publisher Division
New York Chicago San Francisco London Madrid Mexico City Milan New Delhi
San Juan Singapore Sedney Toronto , 2003.

11. Gerhard Krauss . Biochemistry of Signal Transduction and Regulation .
Second Edition . Wiley —VCH Verlag GmbH, 2001.

12. Robert L. Souhan (Editor) ; Ian Tannock (Editor) ; Petter Hohenberen (Editor) ; Jean-Claude Horiot (Editor) . Oxford Textbook of Oncology .
Oxford Press , 2002.

13. Lehninger . Principle of Biochemistry .
Fourth Edition . 2005. Acrobat_60.

14. Mathews , Van holde , Ahern . Biochemistry.
Third Editon . Web site , 2004.

15.Robert M. Watcher (Editor) ; lee Golman (Editor) ; Harry Hollander (Editor) . Hospital Medicine .
Lippincott Williams & Wilkin , 2000.

16. Bernard N. Fields ; David M. Knipe ; Teter M. Howley . Fundamental Virology .
Lippincott Williams & Wilkin .
Philadelphia — Baltimore — New York —London — Boenos Aires — Hong kong — Sydney — Tokyo , 1996.

17. Voet Pratt . Biochemistry . Web site , 2005.

18. Walter R. Wilson ; W. Laurence ; MD Drew ; Nancy K., Phd hery ; Merle A., MD Sande ; David A., Md Relman ; James M.., MD Steckelberg ; Julie Louise ; MD Gerberding . Current Diagnosis & Treatment Infectious Diseases 1th Editon .
Mc Graw-hill/ Appleton & Lange , 2001.

19. Prein Seth . Adenovirus : Basic Biology to Gene Therapy .
R . G . Lands Company , Austin , Texas USA , 1999 .

Một số trang Web

http://www.nature.com/gt/journal/v12/n14/3302503a.html.

http://www.medicalnewstoday.com/medicalnews.php?newid=27707.

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

www.ebi.ac.uk/databases.

www.nig.ac.jp/section/service.html.

www.expassy.org.

www.atcc.org.

www.dsmz.de