

# ライフサイエンスにおける 統合データベースの活用

今西 規

産業技術総合研究所 バイオメディシナル情報研究センター  
分子システム情報統合チーム

# どんなデータベースを研究で使っていますか？

「標準的」なデータベースとしては、以下があります。

- DNA塩基配列 - - - - DDBJ/EMBL/GenBank
- タンパク質配列 - - - - UniProt
- タンパク質立体構造 - - PDBj
- 遺伝子発現 - - - - - GEO (NCBI)
- パスウェイ - - - - - KEGG
- ゲノム多型 - - - - - dbSNP (NCBI)
- 文献情報 - - - - - PubMed (NCBI)

でも、それだけで満足ですか？

## 「発見競争」に勝つには？

- 生物学は大量の情報が利用できる状況にあり、  
"data-driven science"が盛んである。
- 標準的なデータベースは誰でも使っている。
- 他の人と同じデータベースを使っているにも、「発見競争」に勝てないかもしれない。
- ではどうすればいいか...

# 解決策

1. その道のプロに聞け。
2. あらゆる手段で情報を集めよ。  
ポータルサイト・リンク集  
書籍・論文集  
検索の活用

# ポータルサイト・リンク集

# Nucleic Acids Research, Database Categories List

The screenshot shows the Oxford Journals website for Nucleic Acids Research. The page features a navigation bar at the top with links for 'CONTACT US', 'MY BASKET', and 'MY ACCOUNT'. Below this is a large header with the journal title 'Nucleic Acids Research'. A secondary navigation bar includes 'ABOUT THIS JOURNAL', 'CONTACT THIS JOURNAL', 'SUBSCRIPTIONS', 'CURRENT ISSUE', 'ARCHIVE', and 'SEARCH'. A breadcrumb trail reads: 'Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper Categories'. The main content area is titled 'NAR Database Categories List' and lists various database types: Nucleotide Sequence Databases, RNA sequence databases, Protein sequence databases, Structure Databases, Genomics Databases (non-vertebrate), Metabolic and Signaling Pathways, Human and other Vertebrate Genomes, Human Genes and Diseases, Microarray Data and other Gene Expression Databases, Proteomics Resources, Other Molecular Biology Databases, Organelle databases, Plant databases, and Immunological databases. Two sidebars, one on the right and one at the bottom left, each contain a list of navigation options: 'Compilation Paper', 'Category List', 'Alphabetical List', 'Category/Paper List', and 'Search Summary Papers'. A disclaimer states: 'Oxford University Press is not responsible for the content of external internet sites'. At the bottom, it provides ISSN information (Online ISSN 1362-4962, Print ISSN 0305-1048), copyright information (© 2008 Oxford Journals), and the Oxford Journals logo. Footer links include 'Site Map', 'Privacy Policy', and 'Frequently Asked Questions'. A note at the bottom right says 'Other Oxford University Press sites:'.

<http://nar.oupjournals.org/>

# NAR, Database Listに掲載されている データベースの件数

• Nucleotide Sequence Databases	129
• RNA sequence databases	63
• Protein sequence databases	179
• Structure Databases	131
• Genomics Databases (non-vertebrate)	252
• Metabolic and Signaling Pathways	106
• Human and other Vertebrate Genomes	110
• Human Genes and Diseases	114
• Microarray Data and other Gene Expression Databases	65
• Proteomics Resources	18
• Other Molecular Biology Databases	41
• Organelle databases	26
• Plant databases	100
• Immunological databases	27
• Total	1361

# 文部科学省統合データベースセンター DBCCLS (http://lifesciencedb.jp)

収録データベース: 665件

Search:

Tree List

タイプ: 分類

- [-] トップ(612)
- [-] DB型(343)
- [-] 対象(318)
  - [-] アノテーション(24)
    - タンパク質, アミノ酸特...
    - タンパク質, プロテオ...
    - タンパク質, モチーフ(4)
    - タンパク質, 変異(1)
    - タンパク質, 局在(1)
    - タンパク質, 抗体(2)
    - タンパク質, 機能(9)
    - タンパク質, 立体構造(2)
    - タンパク質, 配列(10)
  - [-] マーカー分子(1)
  - [-] 代謝産物(2)
  - [-] 化合物(1)
  - [-] 医学(2)
  - [-] 形態(8)
  - [-] 文献(6)
  - [-] 比較ゲノム(3)
  - [-] 画像(10)
  - [-] 病気, ガン(4)
  - [-] 相互作用(22)

[データベース カタログの一括ダウンロード](#)

画像を選択してください

表示形式: サムネイル ソート: 無し

CYORF (Cy...  
シアノバクテリア遺...

EICO DB  
マウスのインプリン...

Ensembl  
ゲノム自動アノテーション

FANTOM  
マウス全長cDNA...

GadFly  
ショウジョウバエ...

H-Invitat...  
ヒト遺伝子アノテ...

HA1

KATANA

Kaiko Gen

評価 コメント(0)

LsdbID:  
11360

タイトル:  
CYORF (Cyanobacteria Gene Annotation Database)

URL:  
<http://cyano.genome.jp>

LsdbName:  
シアノバクテリア遺伝子アノテーションワークベンチ

分類  
DB型: プロジェクト  
生物種: 微生物, シアノバクテリア  
対象: アノテーション

カタログ:  
文科省  
wing

組織  
国立大学法人: 京都大学: 化研

説明  
シアノバクテリア研究コミュニティが遺伝子アノテーションを行うためのワークベンチ。一般ユーザはデータの検索、参照、一括ダウンロードが可能。

# 経済産業省統合データベース・ポータルサイト MEDALS

<http://medals.jp/>

MEDALS  
METI database portal for life science

ホーム データベース ツール ダウンロード 総合DB MEDALSツール

English 使い方ビデオ

検索

MEDALSの使い方ビデオを公開しました。[09.03.31]

## 経済産業省ライフサイエンス統合データベースポータルサイト

このウェブサイトは経済産業省統合データベースプロジェクトのポータルサイトです。'MEDALS'は、METI Database portal for Life Scienceの略であり、METIとは経済産業省 (Ministry of Economy, Trade and Industry) の意味です。これまで経済産業省関連機関により、様々なライフサイエンス分野の研究開発プロジェクトが実施されてきました。それらの成果物が効果的に利用されることを目的に、1989年以降の成果物 (データベース、ツール等) の情報を整理・提供していきます。今後も内容を更新・追加していきます。 [もっと見る](#)

 <b>データベース便覧</b> 全リスト (40件) ・様々なデータベースの情報があります <a href="#">DNA・ゲノム</a> (17件) <a href="#">RNA</a> (14件) <a href="#">蛋白質</a> (22件) <a href="#">その他</a> (22件)	 <b>統合データベース</b> ・各種のデータを統合したDBです <a href="#">ヒト遺伝子統合データベース (HInvDB)</a> <a href="#">文芸省統合データベースプロジェクト (連携先)</a> 
 <b>ツール便覧</b> 全リスト (32件) ・データ解析に役立つツールの情報があります <a href="#">DNA・ゲノム</a> (10件) <a href="#">RNA</a> (13件) <a href="#">蛋白質</a> (12件) <a href="#">その他</a> (7件)	 <b>サイト内検索</b> ・MEDALSサイト内のキーワード検索です <a href="#">サイト内の詳細な検索</a>
 <b>ダウンロードページ</b> 全リスト (2件) ・成果物をダウンロードできます <a href="#">ツールダウンロード</a>	 <b>MEDALSツール</b> 全リスト (3件) ・MEDALSで開発した研究ツールです <a href="#">リンク自動管理システム</a> <a href="#">ID一括登録システム</a> <a href="#">新刊関連文献お知らせツール (PubMedScan)</a>

MEDALSについて | お問い合わせ

データベース便覧

ツール便覧

収録: 82件

# 書籍・論文集

# 実験医学増刊

## バイオデータベースとソフトウェア最前線



### 実験医学増刊

生命研究への応用と開発が進む

バイオデータベースとソフトウェア最前線

DNA解析からRNA・タンパク質の機能解明

、エピジェネティクス研究、システム生

物学と誰もが使う文献検索の新機能

\* 森下 真一 / 編 阿久津 達也 / 編

\* 定価 5400円 + 税

\* 2008/04発行

\* ISBN 9784758102902

# 検索の活用

# Google検索

ウェブ 画像 動画 地図 ニュース グループ Gmail その他 ▼ ログイン



"RNA polymerase II"

検索

[検索オプション](#)  
[表示設定](#)

ウェブ全体から検索 日本語のページを検索

ウェブ [+ 検索ツールを表示](#)

"RNA polymerase II" の検索結果 約 603,000 件中 1 - 10 件目 (0.35 秒)

[RNA polymerase II - Wikipedia, the free encyclopedia](#) - [ [このページを訳す](#) ]

**RNA polymerase II** (also called RNAP II and Pol II) is an enzyme found in eukaryotic cells. It catalyzes the transcription of DNA to synthesize precursors of mRNA and most snRNA and microRNA. A 550 kDa complex of 12 subunits, RNAP II is ...

[Stages of transcription - Initiation - RNA polymerase control by ...](#)

[en.wikipedia.org/wiki/RNA\\_polymerase\\_II](http://en.wikipedia.org/wiki/RNA_polymerase_II) - キャッシュ - 類似ページ

[PDF] [RNAポリメラーゼIIと基本転写因子の相互作用 保母 順造](#)

ファイルタイプ: PDF/Adobe Acrobat

Human **RNA polymerase II** subunit hRPB14 is homologous to yeast RNA polymerase I, II, and III subunits(AC19 and RPB11) and is similar to ... transcription initiation by **RNA polymerase II** in soluble extract from isolated mammalian nuclei. ...

[www.saitama-med.ac.jp/jsms/vol28/04/t85\\_95.pdf](http://www.saitama-med.ac.jp/jsms/vol28/04/t85_95.pdf) - 類似ページ

保母順造 著 - 2001 - [関連記事](#)

[PDF] [RNAポリメラーゼIIの構造と機能メカニズム RNA polymerase II ...](#)

ファイルタイプ: PDF/Adobe Acrobat

**RNA polymerase II**: Structure and functional mechanism. Patrick Cramer. (University of Munich, Institute of Biochemistry and Gene Center) . のrNTPのリン酸にOH基が求核反応にて結合するのを促進する。また、もう一方のイオンは次に入って ...

[176\\_s.pdf](#) - 類似ページ

RNA polymerase II  
で約60万件のヒット

# NCBI, Entrezによる[All Databases]検索

The screenshot displays the NCBI Entrez search engine interface. At the top, the NCBI logo and the Entrez logo are visible, along with the text "Entrez, The Life Sciences Search Engine". Below the logo, there are navigation tabs for "HOME", "SEARCH", "SITE MAP", "PubMed", "All Databases", "Human Genome", "GenBank", "Map Viewer", and "BLAST". The search bar contains the query "RNA polymerase II" and has buttons for "GO", "Clear", and "Help".

Below the search bar, a message states: "Result counts displayed in gray indicate one or more terms not found". The search results are organized into two columns of database entries, each with a result count in a gray box, a database icon, and a description:

Database	Result Count	Description
PubMed	8054	biomedical literature citations and abstracts
PubMed Central	13370	free, full text journal articles
Site Search	10	NCBI web and FTP sites
Books	264	online books
OMIM	271	online Mendelian Inheritance in Man
OMIA	1	online Mendelian Inheritance in Animals
Nucleotide	17217	Core subset of nucleotide sequence records
EST	6233	Expressed Sequence Tag records
GSS	168	Genome Survey Sequence records
Protein	22110	sequence database
Genome	192	whole genome sequences
Structure	132	three-dimensional macromolecular structures
Taxonomy	none	organisms in GenBank
SNP	none	single nucleotide polymorphism
dbGaP	1	genotype and phenotype
UniGene	1641	gene-oriented clusters of transcript sequences
CDD	90	conserved protein domain database
3D Domains	2755	domains from Entrez Structure
UniSTS	235	markers and mapping data
PopSet	112	population study data sets
GEO Profiles	18670	expression and molecular abundance profiles
GEO DataSets	102	experimental sets of GEO data
Cancer Chromosomes	none	cytogenetic databases
PubChem BioAssay	7	bioactivity screens of chemical substances
PubChem Compound	15	unique small molecule chemical structures

RNA polymerase II  
で約8万件のヒット

# Bioinformatics Harvesterによる検索

<http://harvester.fzk.de/harvester/>



Karlsruhe Institute of Technology

Harvester crawls and crosslinks the following bioinformatic sites:

[4DXp](#) - [AceView](#) - [BLAST](#) - [Biocompare](#) - [CDART](#) - [CDD](#) - [ensEMBL](#) - [Entrez](#) - [FishMap](#) - [Galaxy](#) - [UCSC GenomeBrowser](#) - [gfp-cDNA](#) - [Google Scholar](#) - [gpubmed](#) - [Harvester2](#) - [H-lay](#) - [HomoloGene](#) - [iHOP](#) - [IPI](#) - [MapView](#) - [MGI](#) - [MINT](#) - [Mitochcek](#) - [OMIM](#) - [PolyMeta](#) - [PSORT II](#) - [RGD](#) - [SMART](#) - [SOSUI](#) - [STRING](#) - [TAIR](#) - [Unigene](#) - [UniprotKB](#) - [Wikipedia](#) - [WikiProtein](#)  
[Harvester Sequence Search device](#) - [YaCy-Sciencenet p2p search engine](#)  
See our [H-Wiki](#) for latest activities... Have fun...

## Bioinformatic Harvester IV (beta)

...serving 10.000s of pages every day - Note that '\*' and '?' wildcards are supported.

## Search the human Harvester

Enter one or more keywords to search for. Note that '\*' and '?' wildcards are supported.

Search for:   Results per page:

Match:  any search words  all search words

Search results for: "rna polymerase II"

Your search query contained too many common words to return the entire set of results available. Please try again with a more specific query for better results.

More than 200 results found

21 pages of results.

1. [human protein IPH0007251 - POLR2D, DNA-directed RNA polymerase II subunit RPB4 - Bioinformatic Harvester @ KIT](#)

2. [human protein IPH0003311 - POLR2L, DNA-directed RNA polymerase I, II, and III subunit RPSDC5 - Bioinformatic Harvester @ KIT](#)

3. [human protein IPH0016258 - POLR2C, DNA-directed RNA polymerase II subunit RPB3 - Bioinformatic Harvester @ KIT](#)

4. [human protein IPH0013677 - MED21, Mediator of RNA polymerase II transcription subunit 21 - Bioinformatic Harvester @ KIT](#)

5. [human protein IPH0003308 - POLR2H, DNA-directed RNA polymerase I, II, and III subunit RPARC3 - Bioinformatic Harvester @ KIT](#)

RNA polymerase II  
で210件のヒット

# DBCLSによるライフサイエンスデータベース横断検索

<http://lifesciencedb.jp/>

The screenshot shows the DBCLS search interface. The search term 'RNA polymerase II' is entered in the search bar. The left sidebar shows a tree view of databases, with '全てのデータベース (36108)' selected. The main content area displays search results for 'RNA polymerase II'. The first result is 'Het-PDB Navi. -PDB Code Search [het-pdbnavi]' with a search result URL. The second result is 'Het-PDB Navi. -PDB Code Search [het-pdbnavi]' with a search result URL. The bottom section shows a list of related genes/proteins, including 'DNA-directed RNA polymerase II 140 kDa polypeptide' and 'DNA-directed RNA polymerase II 33 kDa polypeptide'. The page number is 1 of 690 for the first result and 1 of 1 for the second result.

生命科学データベース横断検索 RNA polymerase II データベース一覧 | HELP | 検索結果URL表示

データベース

- 全てのデータベース (36108)
- 統合DBプロジェクト (0)
- 文献 (291)
- 蛋白質核酸酵素 (15)
- 蛋白質核酸酵素 (近刊) (21)
- 文科学「ゲノム特定領域」年次報告書 (1)
- 文科学「旧ゲノム特定領域」最終報告書 (14)
- 日本生物物理学会大会発表要旨 (4)
- BIOPHYS: 生物物理 (4)
- 科学新聞記事- (0)
- 特許関連文書 (6448)
- 日本国特許 (6448)
- 用語解説 (75)
- 日本語版ウィキペディア (35)
- Wikipedia (English) (40)
- ゲノム・遺伝子・RNA (5247)
- 遺伝子発現・転写制御 (1825)
- タンパク (5959)
- 酵素・相互作用・立体構造 (626)

検索結果 ALL

Het-PDB Navi. -PDB Code Search [het-pdbnavi]  
Search PDB Code PDB Code Hetero-Atom Code Protein Name  
Hetero-Atom Name Last update: 31 Oct 2008 UM PDBsum : 2VUM  
ALPHA-AMANITIN INHIBITED COMPLETE RNA POLYMERASE II  
ELONGATION COMPLEX PDB Data Summary. ANSFERASE Title:  
ALPHA-AMANITIN INHIBITED COMPLETE RNA POLYMERASE II  
ELONGATION COMPLEX Compound: MOL\_ID: 1; MOLECULE:  
DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB1; CHAIN: A;  
SYNONYM: RNA POLYMERASE II SUBUNIT B1, RNA POLYMERASE II  
SUBUNIT 1, DNA-DIREC  
[http://hetpdbnavi.nagahama-i-bio.ac.jp/create\\_search\\_result\\_page.php](http://hetpdbnavi.nagahama-i-bio.ac.jp/create_search_result_page.php)

Het-PDB Navi. -PDB Code Search [het-pdbnavi]  
Search PDB Code PDB Code Hetero-Atom Code Protein Name  
Hetero-Atom Name Last update: 31 Oct 2008 3CQZ PDBsum : 3CQZ  
CRYSTAL STRUCTURE OF 10 SUBUNIT RNA POLYMERASE II IN  
COMPLEX WITH THE INHIBITOR ALPHA- ANSCRIPTION Title: CRYSTAL  
STRUCTURE OF 10 SUBUNIT RNA POLYMERASE II IN COMPLEX WITH  
THE INHIBITOR ALPHA- IN Compound: MOL\_ID: 1; MOLECULE:

Page 1 of 690 検索結果34495件中 1 - 50件目を表示

遺伝子リスト -> Human

DNA-directed RNA polymerase II 140 kDa  
polypeptide [GENE:5431][GNP:5431][UNP:P30876][HGNC:9168]  
elongation factor RNA polymerase II-like  
3 [GENE:80237][GNP:80237][UNP:Q9HB65][HGNC:23113]  
RNA polymerase II elongation factor  
ELL2 [GENE:22935][GNP:22935][UNP:Q00472][HGNC:17064]  
DNA-directed RNA polymerase II 33 kDa  
polypeptide [GENE:5432][GNP:5432][UNP:P19387][HGNC:9169]  
DNA-directed RNA polymerase II 19 kDa  
polypeptide [GENE:5436][GNP:5436][UNP:P62487][HGNC:9154]  
DNA-directed RNA polymerase II 16 kDa

Page 1 of 1 検索結果20件中 1 - 20件目を表示

RNA polymerase II  
で36,108件のヒット

# リンク自動管理システムによるIDを使った横断検索

<http://biodb.jp/>

(対象はヒトのみ)

*Hyperlink management system*

HUGO gene symbol POLR2A [All Databases] Search

H-invDB (Transcript view)

POLR2A	HIT000194797	<a href="http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000194797">http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000194797</a>
POLR2A	HIT000262540	<a href="http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000262540">http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000262540</a>
POLR2A	HIT000322605	<a href="http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000322605">http://www.h-invitational.jp/hinv/psoup/transcript_view?hit_id=HIT000322605</a>

H-invDB (Locus view)

POLR2A	HIX0039202	<a href="http://www.h-invitational.jp/hinv/psoup/locus_view?hit_id=HIX0039202">http://www.h-invitational.jp/hinv/psoup/locus_view?hit_id=HIX0039202</a>
--------	------------	---

H-invDB (G-integro)

POLR2A	HIX0039202	<a href="http://www.h-invitational.jp/hinv/g-integro/cgi-bin/f_genemap.cgi?hit_id=HIX0039202">http://www.h-invitational.jp/hinv/g-integro/cgi-bin/f_genemap.cgi?hit_id=HIX0039202</a>
--------	------------	---

H-invDB (PPI view)

POLR2A	HIP000015612	<a href="http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000015612">http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000015612</a>
POLR2A	HIP000062338	<a href="http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000062338">http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000062338</a>
POLR2A	HIP000062339	<a href="http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000062339">http://www.h-invitational.jp/hinv/ppi/ppi_view.cgi?hip=HIP000062339</a>

NCBI (Nucleotide)

POLR2A	NM_000937	<a href="http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&amp;val=NM_000937">http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&amp;val=NM_000937</a>
--------	-----------	---

NCBI (Entrez Gene)

POLR2A	5430	<a href="http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&amp;cmd=Retrieve&amp;dopt=full_report&amp;list_uids=5430">http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&amp;cmd=Retrieve&amp;dopt=full_report&amp;list_uids=5430</a>
POLR2A	POLR2A	<a href="http://www.genenames.org/data/hgnc_data.php?match=POLR2A">http://www.genenames.org/data/hgnc_data.php?match=POLR2A</a>

POLR2A  
で20件のヒット

# ヒト遺伝子と転写産物に関する 統合データベース H-InvDB

# H-Invitational プロジェクト

ミレニアム  
ゲノム  
プロジェクト

## Human Full-length cDNA Annotation **Invitational** (H-Invitational)

世界のヒト完全長cDNAクローン情報を収集し、包括的なアノテーション(注釈付け)を行うことにより、トランスクリプトームの統合データベース構築を目標とする。

生物情報解析研究センター(JBIC、産総研)およびDDBJ(遺伝研)が主催し、世界の44研究機関からの120人以上の研究者が参加した。

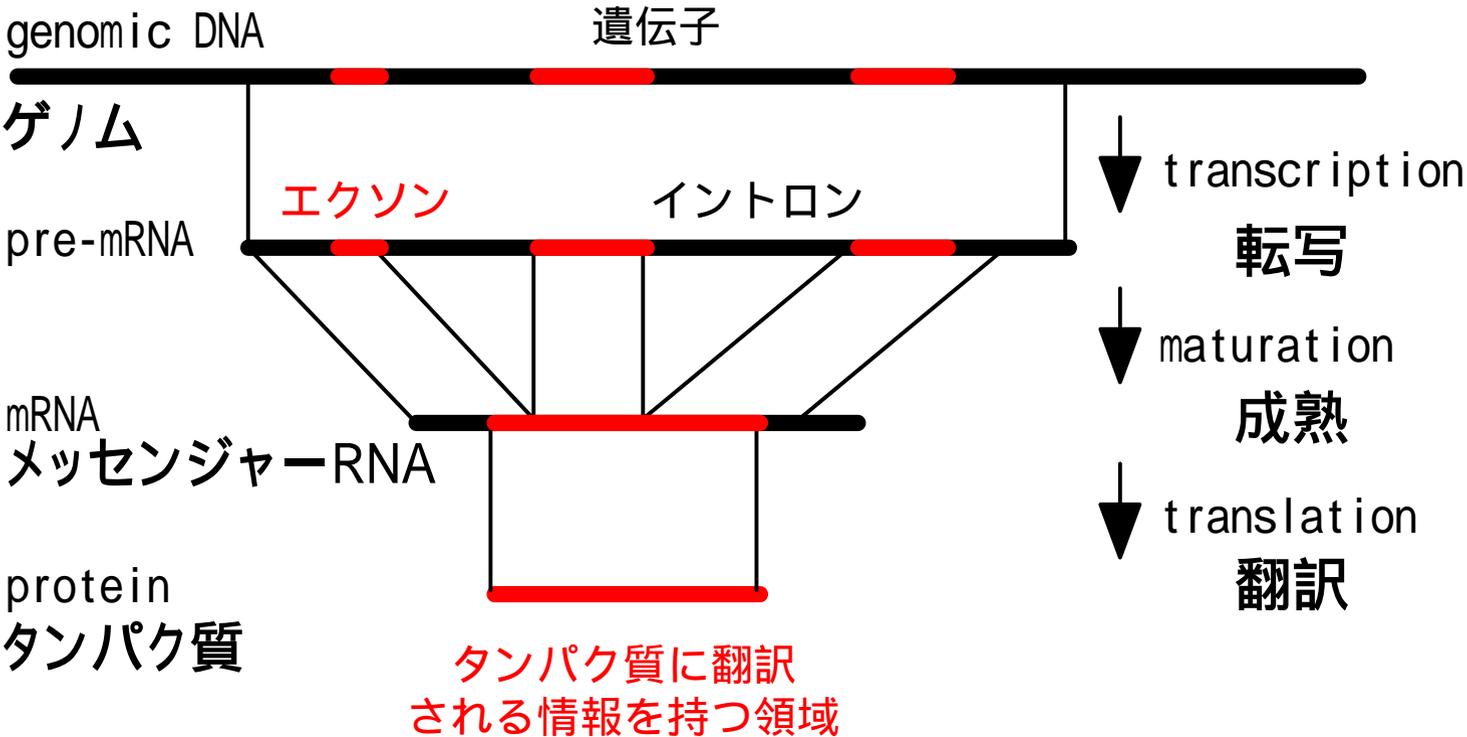


H-Invジャンボリー風景  
(2002年8月)



統合データベース: H-InvDB  
2004年4月公開

# 遺伝情報の伝達様式



# 完全長cDNA (FLcDNA)

- 全長の成熟mRNAからクローニングされる。
- FLcDNAsはイントロンを含まない。
- Open reading frames (ORFs)の予測が比較的容易である。
- 選択的スプライシングバリエーションに対応するFLcDNAが入手可能である。

# H-InvDBにおけるアノテーションの概要

大量の転写産物の配列データを対象に、

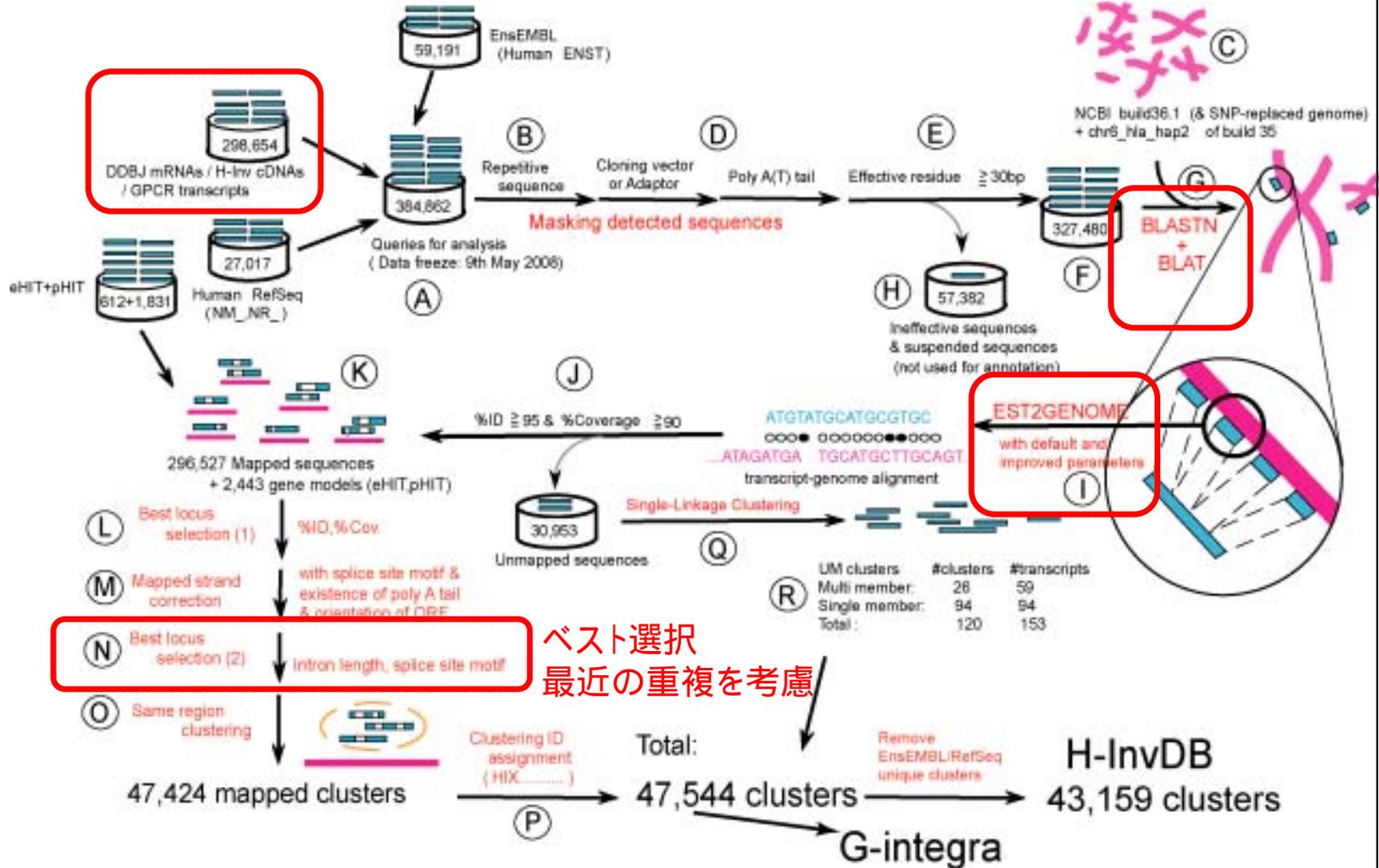
(1) 24項目にわたる大規模な計算機解析と

(2) 20名以上のアノテータ・研究者による文献確認・データ精査、  
の両方を実施して、データベースに格納する。



# 転写物配列のゲノム配列へのマッピングとクラスタリング

ゲノム配列2種類



ベスト選択  
最近の重複を考慮

# H-InvDB リリース 6.2 (2009年3月公開)

[www.h-invitational.jp](http://www.h-invitational.jp) ([hinv.jp](http://hinv.jp))

ヒトの全遺伝子と転写産物を対象とした統合データベース。合計219,765件の完全長cDNAとDDBJ/EMBL/GenBankに登録されたmRNAの情報を含む。

The screenshot shows the H-InvDB website interface. At the top, there's a search bar and navigation links. Below that, a main content area with a sidebar on the left. A yellow box labeled 'トップページ' (Top Page) points to the main content area. Below the screenshot is a diagram titled 'Main views in H-InvDB' and 'Subdatabases in H-InvDB'. The 'Main views' section shows 'Transcript view' and 'Locus view' with a double-headed arrow between them. The 'Subdatabases' section shows several boxes: 'Evoia' (分子進化データベース), 'G-Integra' (ゲノム統合データベース), 'PPI view' (タンパク質間相互作用情報), 'H-ANGEL' (遺伝子発現プロファイル), 'Gene family/group' (遺伝子ファミリー詳細データベース), and 'DiseaseInfo Viewer' (疾患関連データベース). A yellow box labeled 'データベース構成' (Database Structure) points to this diagram.

## H-InvDBのアノテーション項目

- 遺伝子構造
- スプライシング変異体
- 機能性RNA
- タンパク質の機能
- 機能ドメイン
- 細胞内局在
- 代謝パスウェイ
- タンパク質立体構造
- 疾患との関連
- 遺伝的多型 (SNPs)
- 遺伝子発現パターン
- 分子進化的特徴
- タンパク質間相互作用

# H-InvDBのタンパク質分類(release6.2)

転写産物の総数 (HIT):219,765

クラスターの総数 (HIX):43,146

機能アノテーション・カテゴリー	件数*
<b>protein coding</b>	<b>35,387</b>
<b>I:</b> Identical to known human protein (experimentally validated)	13,337
<b>II:</b> Similar to known protein	3,464
<b>III:</b> InterPro domain-containing protein	2,531
<b>IV:</b> Conserved hypothetical protein	4,100
<b>V:</b> Hypothetical protein	5,490
<b>VI:</b> Hypothetical short protein (20-80 aa)	5,628
<b>VII:</b> Pseudogene candidate (transcribed)	837
<b>non-protein-coding</b>	<b>7,759</b>

機能が精査されたCategory I, II, IIIおよび進化的保存性の高いCategory IVの合計23,432件は最も信頼性の高いヒト遺伝子セットと考えられる。

\*代表配列の件数

# H-InvDB遺伝子リスト特徴抽出ツール

## H-InvDB Enrichment Analysis Tool (HEAT)

統合データベースからの  
知識発見支援ツール

入力された遺伝子セットに  
共通な特徴を探すツール。

解析対象となるアノテーション項目：

染色体バンド、  
遺伝子ファミリー、  
Gene Ontology (GO)、  
機能ドメイン(InterPro)、  
立体構造ドメイン(SCOP)、  
KEGGパスウェイ、  
細胞内局在予測(Wolf PSORT)、  
組織特異的遺伝子発現(H-ANGEL)

実行ボタン →

<http://hinj.jp/heat/> にて公開中。

# 遺伝子リスト特徴抽出ツールの実行例

知識発見支援ツール

投入した遺伝子名： 糖尿病関連遺伝子 6 個

(ADIPOQ, CAPN10, PPARG, TCF7L2, HNF4A, HHEX)

No	アノテーション	出現回数	全代表配列中の出現回数	P 値*
1	<a href="#">GO: 0003677</a>	4	861	0.000004736
2	<a href="#">GO: 0006355</a>	4	1078	0.00001150
3	<a href="#">SCOP: g.39.1.2</a>	2	41	0.00002075
4	HIF0000022	2	44	0.00002378
5	<a href="#">InterPro: IPR001723</a>	2	49	0.00002928
6	<a href="#">GO: 0004879</a>	2	49	0.00002928
7	<a href="#">InterPro: IPR001628</a>	2	51	0.00003164
8	<a href="#">InterPro: IPR000536</a>	2	52	0.00003285
9	<a href="#">SCOP: a.123.1.1</a>	2	52	0.00003285
10	<a href="#">InterPro: IPR008946</a>	2	53	0.00003409

\* Fisherの正確確率  
アノテーション

DNA結合性

転写調節

核受容体

ステロイドホルモン受容体ファミリー (44)

ステロイドホルモン受容体

リガンド依存性の核受容体活性

Zinc-finger, 核ホルモン受容体型

核ホルモン受容体、リガンド結合、コア

核受容体、リガンド結合ドメイン

核ホルモン受容体、リガンド結合