H-InvDB

Annotated Human Gene Database
a comprehensive annotation resource for human genes and transcripts

BIRC Biomedicinal Information Research Center

AIST

Chisato Yamasaki (1), Katsuhiko Murakami (2), Jun-ichi Takeda (1), Yoshiharu Sato (1), Akiko Noda (1), Ryuichi Sakate (1), Takuya Habara (1), Hajime Nakaoka(2,3), Fusano Todokoro (2,4), Akihiro Matsuya (2,5), Tadashi Imanishi (1) and Takashi Gojobori (1, 6)

1: BIRC, AIST, 2: JBIC, 3: C's Lab Co., Ltd., 4: DYNACOM Co., Ltd., 5: Hitachi Ltd., 6: CIB-DDBJ, NIG 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan.

H-Invitational Database (H-InvDB; http://www.h-invitational.jp/) is an integrated database of human genes and transcripts. By extensive analyses of all human transcript, we provide curated annotations of human genes and transcripts that include gene structures, alternative splicing isoforms, non-coding functional RNAs, protein functions, functional domains, sub-cellular localizations, metabolic pathways, protein 3D structure, genetic polymorphisms, relation with diseases, gene expression profiling, molecular evolutionary features, protein-protein interactions (PPIs) and gene families/groups. The latest release of H-InvDB (release 6.0) provide annotation for 219,765 human transcripts in 43,159 human gene clusters based on human FLcDNAs and mRNAs. H-InvDB consists of two main views, the Transcript view and the Locus view, and six auxiliary databases with web-based viewers; G-integra, H-ANGEL, DiseaseInfo Viewer, Evola, PPI view and Gene Family/Group view. We also provides several data mining tools such as "Navi search": consists of 16 search contents each of which includes items for the search condition
URL: http://www.h-invitational.jp/hinv/c-search/hinvNaviTop.jsp "PANDA": Priority ANalysis for Disease Association (PANDA) system
URL: http://www.h-invitational.jp/panda/app
H-InvDB now provides web service APIs of SOAP and REST to use H-InvDB data in programs.
URL: http://www.h-invitational.jp/hinv/hws/doc/

5th Anniversary !! Since 20th April, 2004.

H-InvDB databases

L Locus view

H-InvDB provides annotation items for each HIX (H-Invitational cluster) in Locus view

10 Locus info. Locus information mapping of HIX (H-Invitational cluster) in the human genome; chromosome number, location, strand, chromosome band, disease relationship and links to corresponding RefSeq and Ensembl genes, etc.

11 AS Alternative splicing (AS) information annotation on alternative splicing isoforms.

12 Expression Gene expression information tissue-specific expression in 10 tissue categories determined by iAFLP data.

13 Disease info Disease/pathology information disease related information related to HIX: known disease-related genes and co-localized orphan pathology with the name of the disease and OMIM ID.

L Locus view

T Transcript view

T Transcript view

H-InvDB provides annotation items for each HIT (H-Invitational transcript) in Transcript view

1 Function Gene function information human-curated functional definition, similarity category and related evidences; Gene name; HUGO gene symbols; GO ID; GO term; EC number; EC description; pathway information (KEGG), etc

2 Genome loc. Genome location information mapping of HIT on the human genome; chromosome number, location, strand, chromosome band, and links to corresponding RefSeq and Ensembl genes, etc

3 Transcript info. Transcript information/Transcript quality information transcript length, polyA signal, polyA tail and sequence quality related features.

4 Polymorphism/repeat Polymorphism (SNP, indel), microsatellite (Short Tandem Repeat, STR) and repeat information polymorphism (dbSNP), Microsatellite (Short Tandem Repeat, STR) and repeat information.

5 CDS Predicted CDS information CDS, orientation, codon adaptation index, translation.

6 Motif Motif information location, ID and descriptions of functional motifs (InterPro).

7 Subcellular loc. Subcellular localization information subcellular localization prediction by WolfPSORT, Target P, SOSUI, TMHMM and PTS1.

8 Protein structure Protein structure information (GTOP) assigned PDB and SCOP IDs by reverse PSI-BLAST, and summary prediction of 3D structure by GTOP.

9 Evolutionary info. Evolutionary information Ortholog relationships, phylogenic trees and sequence alignments.

Data mining tools and resources

Navi search
http://h-invitational.jp/hinv/c-search/hinvNaviTop.jsp
[Search Navigation]
Welcome to the H-InvDB Navigation search system.

16 search items
e.g.
location, function, domain, subcellular localization, protein structure, pathway, PPI, gene family, ncRNA, alternative splicing, SNP, ortholgos, expression, etcb

Select search contents

List of genes

PANDA
http://www.h-invitational.jp/panda/app
[Priority ANalysis for Disease Association]

The tool to extract new disease-susceptible genes using H-InvDB, by the concept that the new candidates have similar funtional character to the already reported disease-susceptible genes.

We compared palalog genes, protein function, KEGG pathway,InterPro+Gene Ontology (Molecular Function, Biological Process, Subcellular Localization) between genes.

Select disease

List of candidate genes

Web services
http://h-invitational.jp/hinv/hws/doc/
[SOAP and REST APIs]

H-InvDB Web service provide the various SOAP and REST APIs to use H-InvDB data. Anyone can use the service to develop the application using H-InvDB Web service.

List of APIs

NEW !!

H-InvDB Enrichment Analysis Tool (HEAT),
http://hinv.jp/HEAT/search.php?lang=en

H-InvDB Enrichment Analysis Tool

H-InvDB Enrichment Analysis Tool (HEAT) is a data-mining tool for automatically identifying features specific to a given human gene set. HEAT searches for H-InvDB annotations that are significantly enriched in a user-defined gene set, as compared with the entire H-InvDB representative transcripts. The following features of H-InvDB are analyzed: InterPro, Gene Ontology (GO), KEGG pathway, Chromosomal band, Gene family, SCOP (structural domains), Subcellular localization prediction (by using Wolf PSORT) and Tissue-specific gene expression (10 tissue categories defined in H-ANGEL). This technique is called Gene Set Enrichment Analysis (GSEA), and is popularly used in analyzing results of microarray experiments..

Subdatabases

G-integra [genome mapping view]
G-integra is a genome integrated database, in which we can browse physical maps and gene structures of human, mouse, rat, chimpanzee genome, etc (total 14 species).

H-ANGEL [gene expression view]
H-ANGEL (Human ANatomic Gene Expression Library) is a database of expression patterns that we constructed to obtain a broad outline of expression profiles of human genes. Gene expression data in normal adult human tissues that were generated by using three types of methods and on seven different platforms. The data was collected and categorize into 10 and 40 major tissues

DiseaseInfo Viewer [disease information view]
DiseaseInfo Viewer is a database of known and orphan genetic diseases and their relationships to H-Inv loci withEntrez Gene and OMIM

Evola [evolutional annotation database]
Evola is a database of evolutionary annotation of human genes. It provides sequence alignments and phylogenetic trees of ortholog genes among human and model organisms.

PPI view [Protein-protein interaction (PPI) view]
The PPI view displays H-InvDB human protein-protein interaction (PPI) information.

Gene family/group
H-InvDB Gene Families/Groups are human-curated annotation datasets for the selected gene families/groups.