

ヒト遺伝子の統合データベースH-InvDBと関連データベース・ツール群

H-InvDB, an integrated database of human genes, and related databases and tools

○山崎千里 (1)、武田淳一 (1)、羽原拓哉 (1)、世良実穂 (1)、原雄一郎 (1)、野田彰子 (1)、村上勝彦 (1)、今西規 (1)、五條淑孝 (1,2)
1. 産総研・バイオメディカル情報研究センター、2. 遺伝研・生命情報・DBJ 研究センター
○Chisato Yamasaki (1), Jun-ichi Takeda (1), Takuya Habara (1), Miho Sera (1), Yuichiro Hara (1), Akiko Noda (1), Katsuhiko Murakami (1), Tadashi Imanishi (1), Takashi Gojobori (1, 2)
1: BIRC, AIST; 2: CIB-DBJ, NIG

H-Invitational Database (H-InvDB)はヒトの遺伝子と転写産物を対象とした統合データベースです。

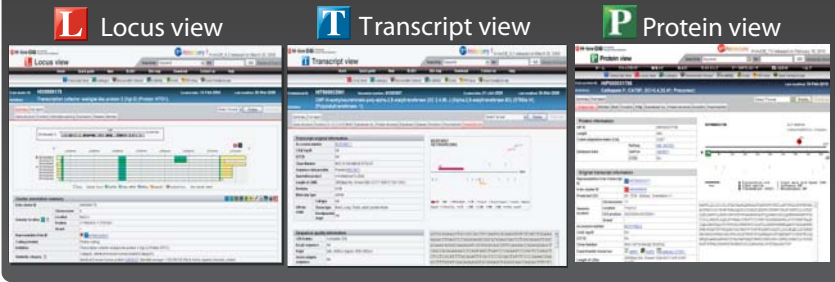
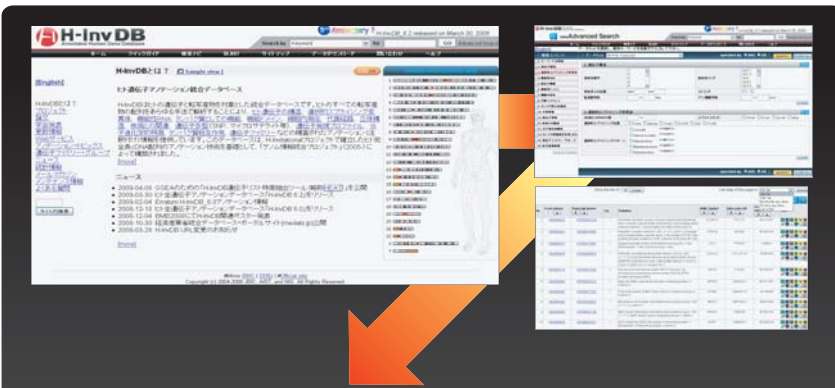
ヒト遺伝子の構造、選択的スプライシング、機能的RNA、タンパク質としての機能、機能ドメイン、細胞内局在、代謝経路、立体構造、疾病との関連、遺伝子多型、遺伝子発現、分子進化的特徴、タンパク質間相互作用、遺伝子ファミリーなどの精密なアノテーション(注釈付け)情報を提供している。最新のH-InvDB release 7.5では、44,886のヒト遺伝子クラスター(遺伝子座)を定義し各種アノテーション情報を付与して<http://www.h-invitational.jp/>より公開しています。

H-InvDBではデータをプログラムから利用することができるwebサービスを公開し、糖鎖遺伝子、糖タンパク質、機能的RNAや多型等有用なヒト分子情報の統合化を行っています。更に、遺伝子構造・機能・発現・進化など16のコンテンツの任意の組み合わせで複合検索ができる検索ナビゲーションシステムや、データマイニングツール(HEAT)、テキストマイニング(LLEGENDA)や自動アノテーションツール(TACT)等の関連データベース・ツール群も提供しており、さまざまな研究ニーズに応えることができます。

H-InvDBは、H-Invitationalプロジェクトで確立したヒト完全長cDNA配列のアノテーション技術を基礎として、ゲノム情報統合プロジェクト(2005-2008)、経済産業省統合データベースプロジェクト(2008-)の主要データベースとして、構築・更新されています。

H-InvDBデータベース

H-InvDB databases



Locus view

- 1 遺伝子座情報 **Locus info** ではゲノム地図、染色体番号、染色体上の位置、ゲノムのストランド、染色体バンドなどのHIXのヒトゲノム上でのマッピングに関するアノテーションデータを提供しています。
- 2 選択的スプライシング変異体情報 **AS** ではHIXに含まれるスプライシング変異体パターンやこの遺伝子座におけるクラスタリングの特徴に関するアノテーションデータを提供しています。
- 3 遺伝子座発現プロファイル **Expression** では10の発現組織カテゴリーに分類されたiAFLP発現データ解析結果に基づき、HIXの遺伝子発現の組織特異性に関するアノテーションデータを提供しています。
- 4 疾患関連情報 **Disease info** では既知の遺伝性疾患関連遺伝子とHIXと同じゲノム領域にマップされた候補遺伝子が判明していない遺伝性疾患(Go-localized orphan pathology)の2種 類の疾患関連情報に対する疾患名およびOMIM IDを提供しています。

Transcript view

- 5 遺伝子機能 **Function** ではH-Invプロジェクトによる遺伝子機能定義および分類、遺伝子名、HUGO 推奨遺伝子名、ジーンオントロジーIDおよび名称(GO)、酵素番号および酵素名(EC)、代謝経路情報(KEGG)、タンパク質相互作用、遺伝子ファミリーなどのHITの遺伝子機能に関するアノテーションデータを提供しています。
- 6 マッピング情報 **Genome loc.** では染色体番、染色体上の位置、ゲノムのストランド、染色体バンドなど、HITのヒトゲノム上でのマッピングに関するアノテーションデータを提供しています。また対応するRefSeq、Ensembl、EntrezGene、GeneCard遺伝子などもリンクしています。
- 7 転写産物アノテーション **Transcript info.** では転写産物に関する精密情報と配列クオリティ情報に関するデータを提供しています。
- 8 多型(SNP)・マイクロサテライト・リピート **Polymorphism/ repeat** は一塩基多型(SNPs)、マイクロサテライト、反復配列などのアノテーションデータを提供しています。
- 9 アミノ酸翻訳領域予測 **CDS** ではアミノ酸翻訳領域(CDS)、アミノ酸配列、コドン使用頻度の指標(CAI)などのHITのアミノ酸翻訳領域に関するアノテーションデータを提供しています。

Protein view

- 10 モチーフ **Motif** では翻訳領域中に予測された機能的モチーフ(InterPro)のID、名称、位置などのHITの機能的モチーフに関するアノテーションデータを提供しています。
- 11 細胞内局在予測 **Subcellular loc.** ではWoLF PSORT、Target P、SOSUI、TMHMM、PTS1の5種のプログラムによる細胞内局在予測結果の概要を表示しています。
- 12 立体構造予測 **Protein structure** ではGTOPICによる立体構造予測結果の概要(reverse PSI-BLAST法によるPDBおよびSCOP ID)を表示しています。
- 13 分子進化的解析 **Evolutionary information** では共通祖先をもつ異種間相同遺伝子(オルソログ)予測、系統樹、自然淘汰(dN/dS)などの分子進化的解析に基づくアノテーションデータを提供しています。

H-InvDBサブデータベース

H-InvDB sub-databases

G-integra



ヒト、マウス、チンパンジー、ラットなどは動物種のゲノム地図と遺伝子構造を示すことができるゲノムブラウザです。遺伝子、EST、遺伝子予測、反復配列等のゲノム上の位置および構造の比較が可能です。

H-ANGEL



ヒト遺伝子の発現プロファイルデータベースです。7つのプラットフォーム、3種の実験手法から得られたヒト成人、健康人の組織の遺伝子発現データを収集、集約ごとに分類し、各HIXの発現プロファイルを提供しています。

DiseaseInfo Viewer



既知疾患遺伝子の情報と染色体バンド上に同座する遺伝子の形質についての情報を提供するデータベースです。各HIXと疾患、MutationViewとの対応を提供しています。

Evola



ヒトとマウスやラットなどのモデル生物について、比較ゲノム解析とH-InvDBの転写産物解析にもとづくオルソログ情報を提供しています。オルソログとパラログからなる遺伝子ファミリーの種間比較情報も提供しています。

PPI view



ヒトのタンパク質間相互作用(PPI: Protein-Protein Interaction)情報のデータベースです。BID、DIP、MINT、IPD、IntActのタンパク質間相互作用情報を統合し、H-InvDBのヒトタンパク質(HIP)と対応させた情報を提供しています。

Gene family / group view



ヒトの遺伝子ファミリー・グループに属する遺伝子を対象にしたアノテーションデータベースです。手動アノテーションで構築された4つの既知遺伝子ファミリー・グループおよび自動アノテーションで定義された予測遺伝子ファミリーについてアノテーション情報を提供しています。

統合DB

H-InvDBサテライトデータベース&ツール

H-InvDB satellite-databases and tools

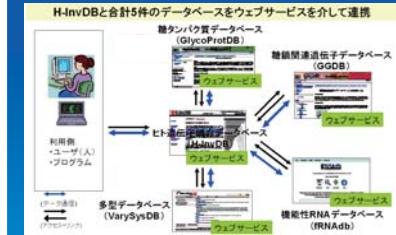
- HEAT (H-InvDB Enrichment Analysis Tool)** H-InvDB遺伝子リスト特徴抽出ツール
- PANDA (Priority ANalysis for Disease Association)** 新規疾病関連遺伝子を探るシステム
- LLEGENDA (Literature-Extracted GENE-Disease Associations)** 文献から機械的に抽出した疾患と遺伝子の関係のデータベース
- H-GOLD (Human-Gene diversity Of Life-style related Disease)** ゲノムワイドな多型マイクロサテライトマーカーのデータベース
- G-compass** 比較ゲノム解析にもとづくヒトとチンパンジー、マウス、ラットなどの保存領域を表示する比較ゲノムブラウザ
- TACT (Transcriptome Auto-annotation Conducting Tool)** 遺伝子の機能を自動予測できる統合的自動アノテーションシステム
- H-DBAS (Human transcriptome DataBase for Alternative Splicing)** 選択的スプライシングバリエーションのデータベース
- VarySysDB** 一塩基多型、挿入欠失多型、STR多型、構造多型、連鎖不平衡領域などのヒト多型情報の統合データベース

<http://www.h-invitational.jp/> [h.inv.jp]

主要論文:
(1) H-InvDB in 2009: extended database and data mining resources for human genes and transcripts. Yamasaki C, et al. (2010) Nucleic Acids Research Jan;38(Database issue):D626-32.
(2) The H-Invitational Database (H-InvDB): a comprehensive annotation resource for human genes and transcripts. Yamasaki C, et al. (2008) Nucleic Acids Research 36, Database issue D793-D799.
(3) Integrative Annotation of 21,037 Human Genes Validated by Full-Length cDNA Clones. T. Imanishi et al. (2004) PLoS Biology 2 (6), 656-675.

ウェブサービスを介したヒト分子情報の統合

経済産業省統合データベースプロジェクト



経済産業省統合データベースプロジェクトでは、H-Invationalを基礎として、糖鎖遺伝子、糖タンパク質、機能的RNAやヒト多型情報等、ヒトの分子情報の統合を進めておりH-InvDBを含む計5件のデータベースでウェブサービスを開発しました。これらのデータベース画面間の連携を進めています。

- 対象データベース:
1. 糖鎖関連遺伝子データベース(GGOB)
2. 糖タンパク質データベース(GlycoProtDB)
3. 機能的RNAデータベース(RNAdb)
4. ヒト多型データベース(VarySysDB)

