大学共同利用機関法人 情報・システム研究機構（ROIS）
ライフサイエンス統合データベースセンター（DBCLS）

統合遺伝子検索 GGRNA

GGRNA: fast and convenient universal search engine for genes and transcript

○内藤雄樹，米澤 明憲，坊農秀雅（情報・システム研究機構 ライフサイエンス統合データベースセンター）
○Yuki Naito, Akinori Yonezawa, Hidemasa Bono (Database Center for Life Science, Research Organization of Information and Systems)

さまざまな切り口で遺伝子を「さがす」または「ながめる」ことができるように，塩基配列，発現情報，立体構造，文献等さまざまな知識を RNA 配列（転写産物）をベースに整理した「RNA 統合データベース」を構築している．その一環として，RefSeq に収録された遺伝子を簡便かつ高速に検索できるウェブサービス GGRNA（http://GGRNA.dbcls.jp/）を構築したので紹介する．GGRNA は，単一の検索窓しかないシンプルなインターフェースから，RefSeq に収録された転写産物の各種 ID，遺伝子名，アノテーション情報，塩基配列およびアミノ酸配列を含む全フィールドを対象に全文検索をおこなうシステムである．とくに塩基配列の検索では，N などの曖昧な塩基を含むパターンや，クエリに対して 2 塩基までのミスマッチを含む配列を高速に検索する．またマイクロアレイのプローブ ID を入力すると，そのプローブの塩基配列をもとに遺伝子を検索することもできる．現在，ヒト，マウス，ラット，ニワトリ，ツメガエル，ゼブラフィッシュ，カタユウレイボヤ，ショウジョウバエ，線虫，イネ，シロイヌナズナ，出芽酵母，分裂酵母に対応している．

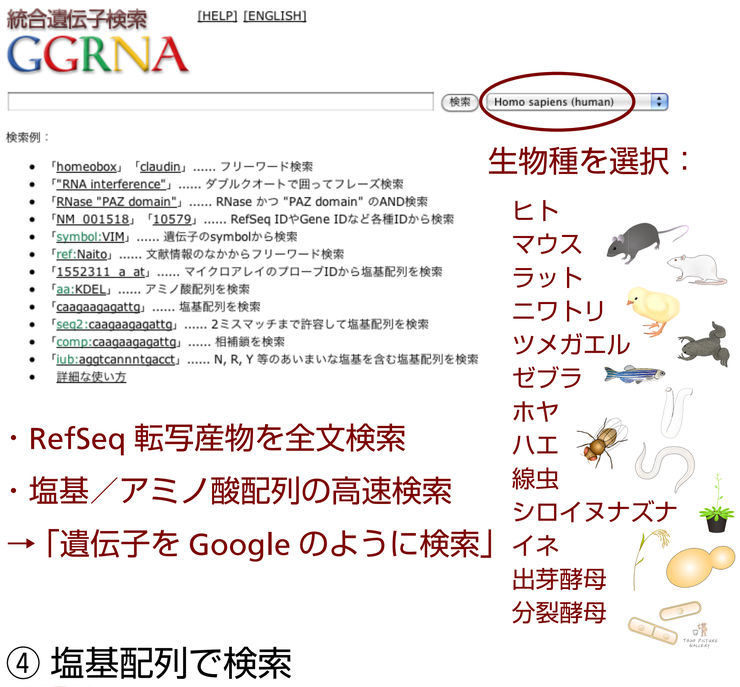
DBCLS における研究開発計画

1. データベースの RDF による統合化
2. 解析プラットフォームによる統合利用環境の整備
3. インターネットを活用した高度検索技術の開発
4. RDF 化に資するオントロジー，辞書，コーパスの整備，標準化
5. 大規模データの利用技術開発
6. 情報統合化・知識発見のためのキュレーション支援
7. 統合データベースに関わるコンテンツの作成，整備

メタ情報による大規模データの整理・再利用促進技術開発
- ダイジェスト版の構築・開発・維持
- RNA 配列を中心とした DB 利用技術開発

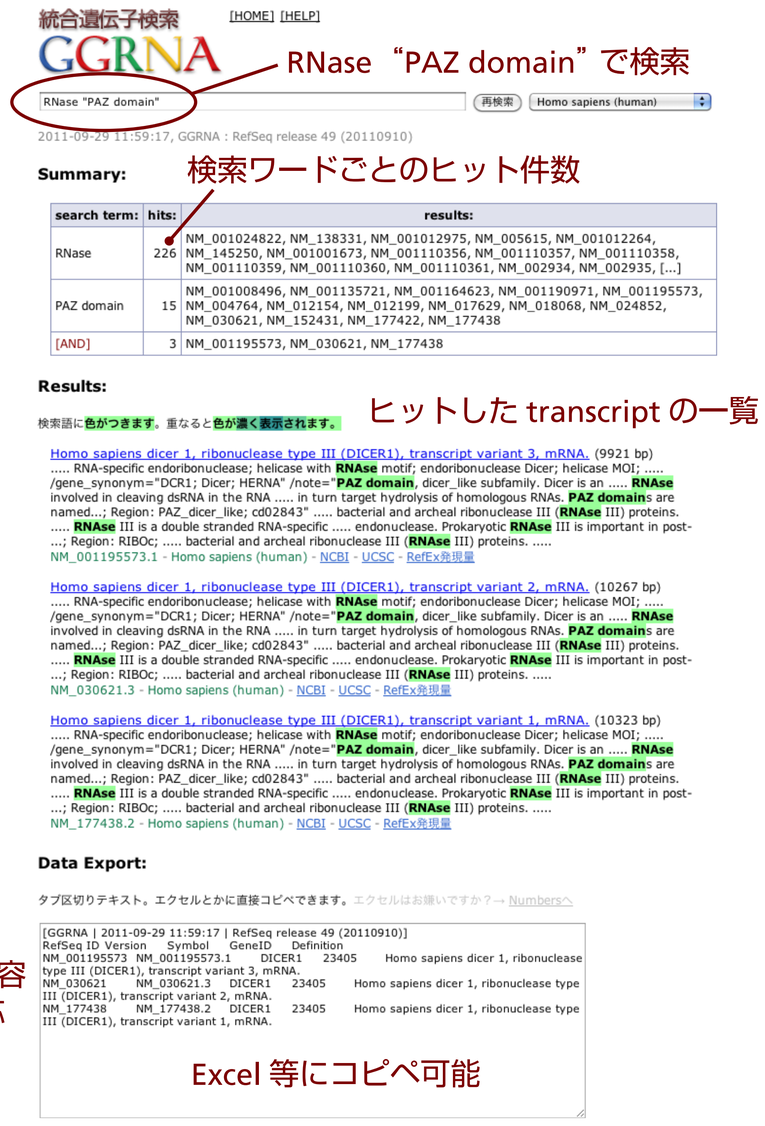
遺伝子発現リファレンスデータセット整備
- RefEx の構築維持管理
- 次世代シーケンサーデータアーカイブ Sequence Read Archive からのデータ抽出技術開発

① トップページ（http://GGRNA.dbcls.jp/）

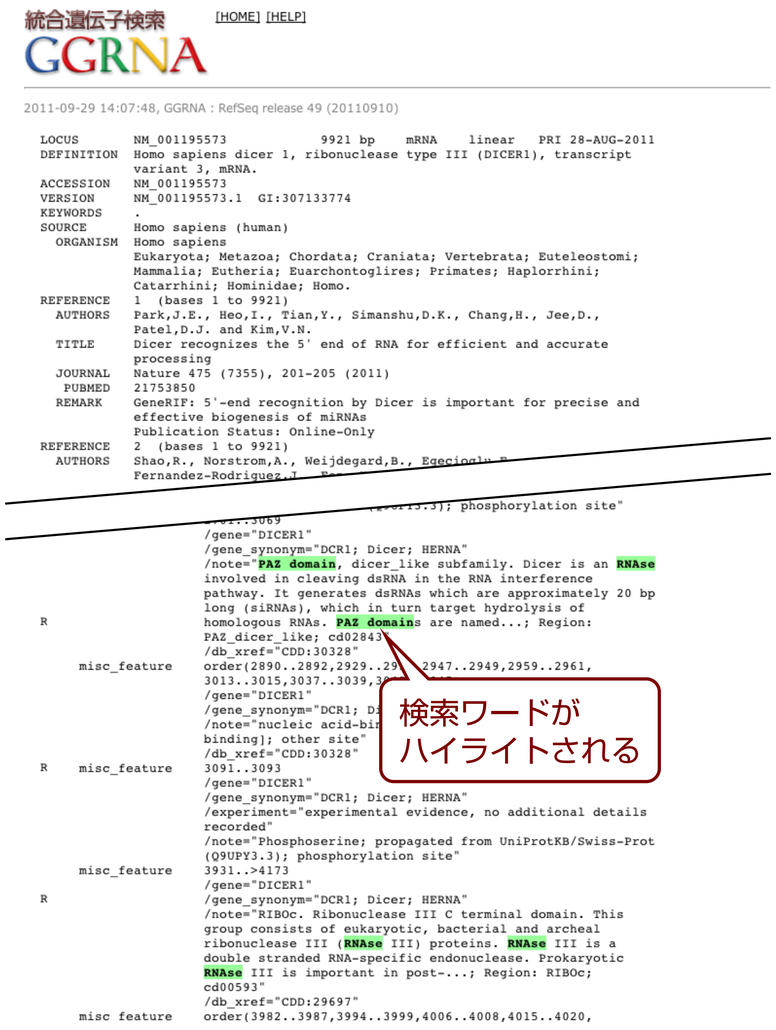


・RefSeq 転写産物を全文検索
・塩基／アミノ酸配列の高速検索
→「遺伝子を Google のように検索」

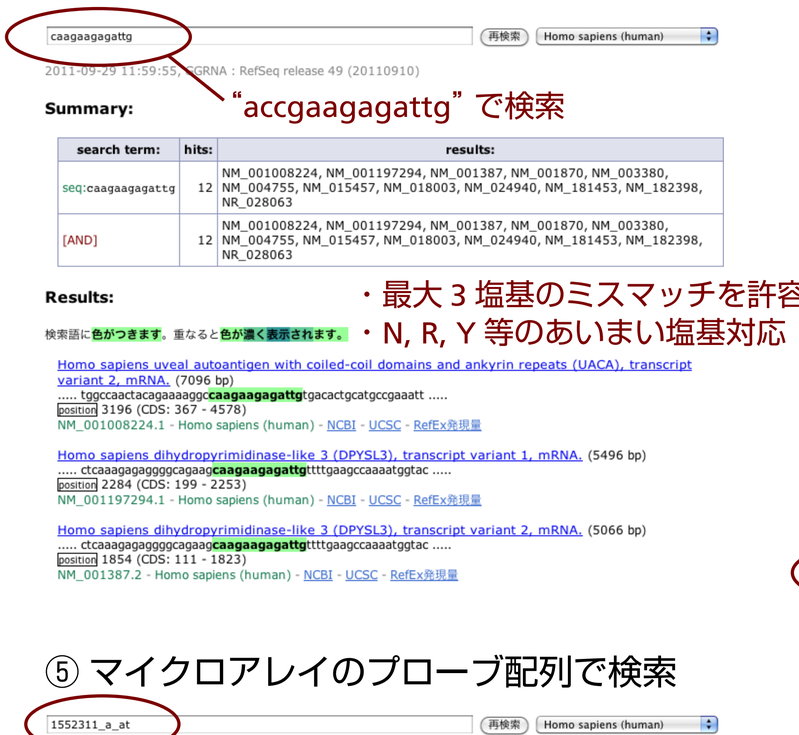
② キーワードで検索



③ 詳細情報表示（RefSeq データ）



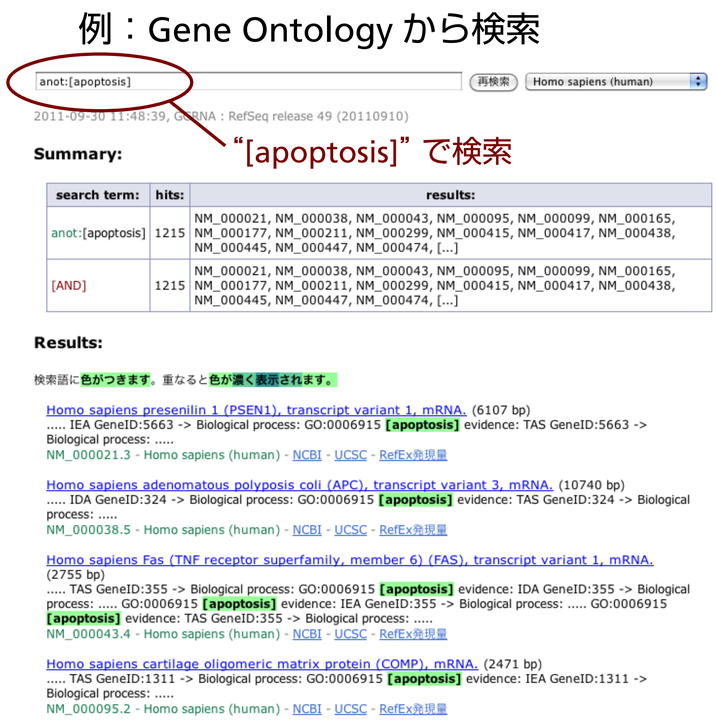
④ 塩基配列で検索



⑤ マイクロアレイのプローブ配列で検索

⑥ アミノ酸配列で検索

⑦ 他のデータベースとの連携
例：Gene Ontology から検索



Gene Ontology や酵素 EC 番号など，RefSeq 以外のデータベースに由来する情報を RefSeq ID と対応付け全文検索できるようにした．

Copyright © 2011 DBCLS licensed by
Creative Commons Attribution 2.1 Japan License.