

共同研究全般にわたるデータベース構築・管理

テーマ4の1の①,②,③,④ (J1)

小テーマ名: (フェーズI) ①マウス総合データベースの基本設計②ハードウェアの仕様決定③ワークフロー管理システム構築

(フェーズI) ①マウス総合データベースの構築②ハードウェアの導入

(フェーズI) ①抗体評価情報管理サブシステムの構築②DNAマイクロアレイ情報管理サブシステムの構築③マウス総合データベースのプロトタイプ構築

(フェーズII) ①抗体評価情報管理サブシステムの改良②DNAアレイ情報管理サブシステムの改良③マウス総合データベースの構築

(フェーズII) ①InGaPデータベースの更新②パスイデータベース InCePの構築③パスイデータベースへのアレイ解析結果の反映

④外部公開データベース GEOからの有用アレイ解析データの効率的取得法の開発

(フェーズII) ①InGaPデータベースの更新②InCePの更新③パスイデータベースへのアレイ解析結果の反映 ④外部公開データベース GEOからの有用アレイ解析データの効率的取得法の開発

(財)千葉県産業振興センター(研究員) 村上 雅利、甲賀弘、川井誠、古賀久芳  
(主任研究員) 古閑比佐志

#### [目的と概要]

かずさDNA研究所が保有するヒト長鎖 cDNA ライブラリー(KIAA)の機能解析研究をすすめるために、モデル生物であるマウス長鎖 cDNA ライブラリーを構築し、遺伝子レベルの機能解析に加え、抗体を用いたタンパク質レベルの機能解析研究を IT 的側面から支援することを目的として研究を行った。

遺伝子レベルの予測データをヒトとマウスを並列する形で取り込み、LIMS(Laboratory Information System)を構築し、以降研究の過程で産出されるデータを一元管理するようにした。

次に、これらのデータをもとに抗原部位を設計する Editor プログラムを統合した。この後の工程も抗体作成・評価のプロセスを管理するワークフローシステムに取り込み、複数担当者で分担する業務を並列して行えるようにした。

mKIAA 遺伝子約 2000 個をプロットしたカスタムアレーを設計支援する仕組みを構築し、LIMS に統合した。さらには、発現データを解析するために正規化などの一連の統計手法を連続して扱うための解析部品を Visual Mining Studio(数理システム)上で開発した。

蓄積した研究成果を外部公開するために InGaP/InCeP を構築した。内部に蓄積した mKIAA に関する知識を遺伝子レベル・mRNA レベル・タンパク質レベルの情報を組織ごとなど階層化して検索できるようにした。加えて、免疫沈降法で得たタンパクタンパク相互作用のデータと文献などの既知の情報とあわせ、mKIAA 機能解析の結果をパスイデに表現して公開した。期待される疾患との関連性からもあわせて検索できるようにした。

ある状況下におけるパスイデの妥当性について検証するために、MAKOT 解析機能を考案した。NCBI が公開する GEO(Gene Expression Omnibus)などの時系列の遺伝子発現データを下に関連の妥当性を時系列にずらした相関についてアニメーション的に見せる仕組みである。

これらの結果、遺伝子機能解析研究のためのパイプラインが完成した。

#### [研究成果の概要と今後の取り組み]

##### (1) マウス総合データベースの構築

ヒトKIAAと相同性のあるマウスmKIAA完全長cDNA配列を格納し、機能予測のための予測プログラムの結果を統合するLIMS(Laboratory Information System)を開発した。バイオインフォマティクス分野においては塩基配列などフラットファイルでデータを持つことが多い。ORACLE等のリレーショナル・データ

ベースをこの分野に適応する際的设计・構築すべきかを検討し適用した。

このシステムはWebの仕組みを採用し、研究所内のネットワークどこからでもアクセスできるように設計したため、レイアウト変更の多いWeb研究者が実験環境にも柔軟に対応することができ、研究業務の効率化を図ることができた。

今後の取り組みとしては研究現場ごとに異なるワークフローを柔軟に定義でき、データ項目の変動にすばやく対応できる開発プラットフォームの研究を進めていきたい。

## (2) 抗体評価情報管理サブシステムの構築

抗体を用いて生体内の遺伝子産物の機能解析をするために、mKIAAのcDNAの配列から抗原部位を選択するEditorプログラムを開発しLIMSに統合した。設計した抗原をもとにウサギに免疫して得た抗体を活用して、生体内のタンパク質の評価実験を行い、結果をWebシステムに集約した（図1-1の上段中央）。

# The Kazusa CREATE Pipeline

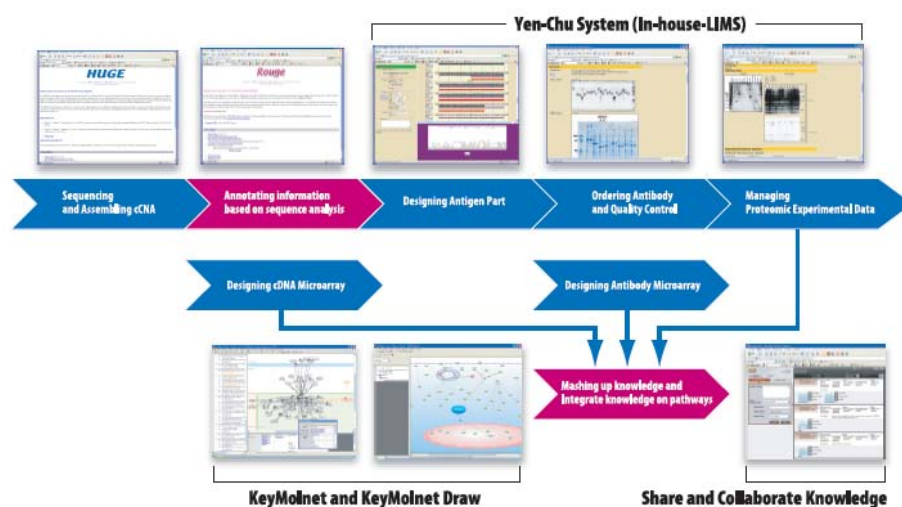


図1-1 地域結集型共同研究で構築したパイプライン

このシステムはヒトとマウスの配列および予測プログラムの結果を並べて表示することにより、ヒトとマウスで交差する領域を抗原部位として選び、ヒトの研究にも転用可能な抗体を設計することに役立った。

今後は後継プロジェクトでこの仕組みはK I A A以外のタンパク質に対する抗体を作成する際に利用される予定である。

## (3) DNAマイクロアレイ情報管理サブシステムの構築

クローンをマイクロアイ割り付ける支援システムおよび解析データを格納するサブシステムを構築した。マイクロアレイで得た発現データは一般にデータ正規化などの取り扱いが難しく、複数の統計手法を組み合わせて行うことが多い。

これらの統計情報の解析結果を解析手法間で受け渡ししながら、一連の処理を自動化するVisual Mining Studio上でいくつかの解析フローを実現した。一般的な統計手法では満足できないものについてはカスタムの統計部品を設計し、実装することで対応した。

今後はエンドユーザーが独自に高度な統計手法を適用する場面は増えてくるものと思われる。目的にあわせた統計手法をコンサルティングする能力が求められる。今回のようなツールで一連の流れを実装したものはそのノウハウを形式知化するものとして、有用性が増すものと思われる。

## (4) InGaPデータベースの構築

遺伝子の配列から予測される情報と抗体評価実験から得られたタンパク質レベルの実験データを遺伝子ごとにカード上にまとめたものがInGaPデータベースを構築し、インターネット上に外部公開した。

# Annotation of mouse KIAA proteins

We have generated more than 1500 mKIAA antibodies and provided several useful information through our web site (<http://www.kazusa.or.jp/create>).

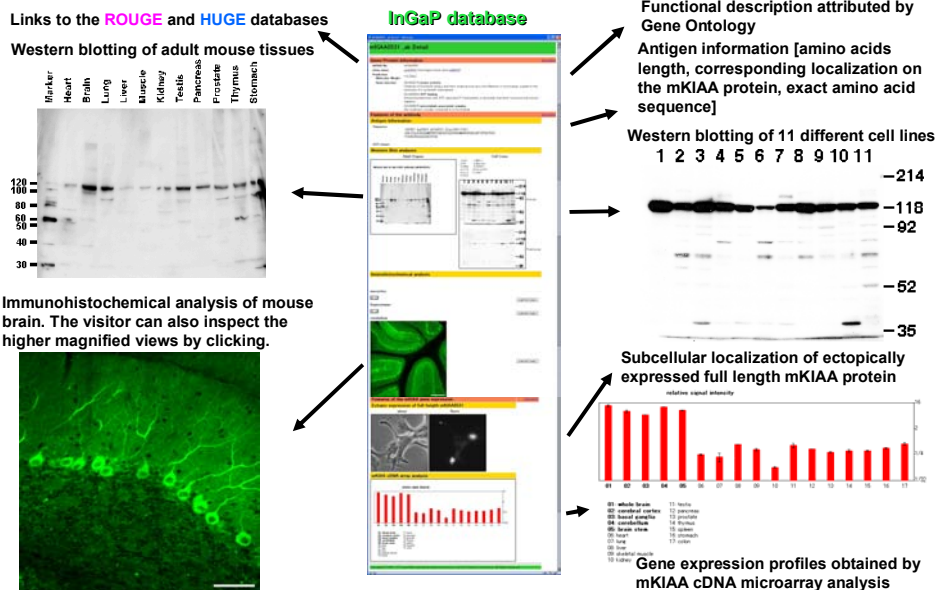


図 1-2 InGaPデータベースで公開したデータの例

2000個のmKIAAについてDNAレベルタンパク質レベルの研究を一つの共同研究体で行い、生データに近い状態で公開することは他に類を見ない。

検索のパターンとしてはmKIAA番号で検索するほか遺伝子レベル、タンパク質レベルの組織別発現により層別して検索することが可能である

## (5) パスウェイデータベースInCePの構築

mKIAA抗体を用いた免疫沈降法によりマウス内在性のタンパク質と相互作用するタンパク質を同定することができた。このデータと遺伝子や関連疾患に関する文献データをあわせて、パスウェイの形にして知識交換する相互接続型プラットフォームを開発し、外部公開した。

## The Concept of the bidirectional pathway knowledge platform

In order to accelerate accumulation of knowledge of the functions of proteins,

- Choose highly crated knowledge by KeyMolnet
- Screen a potential relationship to any disease and drug target around our novel protein-protein interaction
- Exchange a persuasive network diagram, not only a single protein-protein interaction information
- Revise the network diagram and mash up the new knowledge with a freely distributed draw tool
- Attach supplemental data to the network diagram

### Workflow

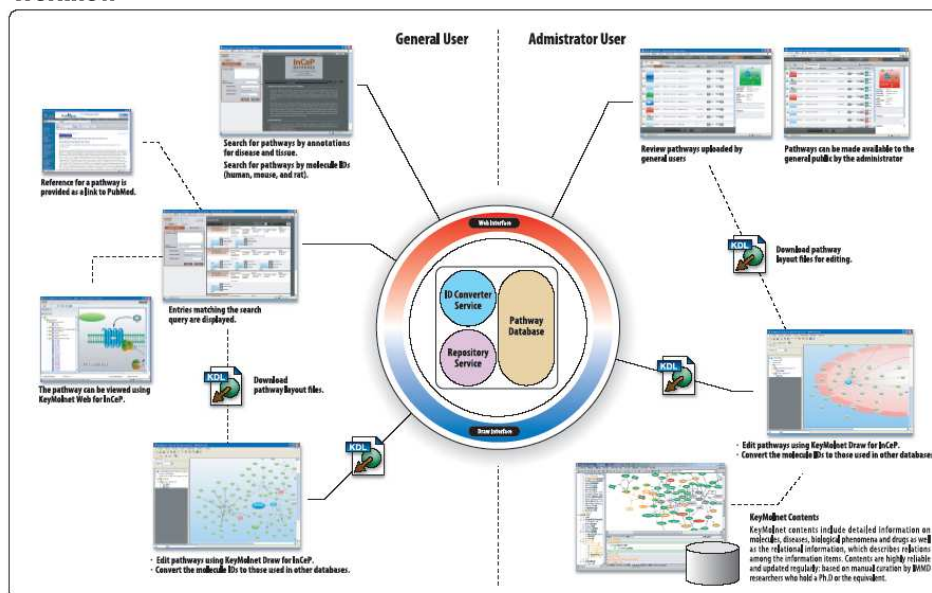


図2-3 KeyMolnet Draw for InCePを用いた相互接続型知識交換プラットフォーム  
現在数十のパスウェイデータを公開している。

このWebサイトで公開しているデータを編集するためのDrawをフリーで配布し、ユーザーサイドからのフィードバックを可能にする仕組みである。従来テキストなどの画像情報のみで交換されていた分子ネットワーク情報を外部データベースへのリンクなどを生かした形で再利用できるようになっている。

また、タンパク質研究では一つのタンパク質が複数の別名を持つことが多いため、研究者の間でも知識の連鎖が起りにくかった。この問題を解決するためにサーバー側にこのDrawツールと通信して、別名および複数の公共データベースのアクセスに変換する仕組みを用意した。この結果、研究者は自分の研究する遺伝子・タンパク質の情報を少ないステップで収集することができるようになった。

KeyMolnetデータベースをあわせて活用することにより、疾患関連性に似着いての情報を効率的に得ることができるようになった。

今後は教育現場で活用してもらいなどこのDrawツールの普及に努めていきたい。

(6) パスウェイデータベースへのアレイ解析結果の反映

GEOなどの一般に公開されている発現データを用いて作成したパスウェイの確からしさを確認するプログラム(MAKOT)を用いた解析結果もInCePシステムで公開した。

フェーズIIIにおける展開のためにKeyMolnet Draw上にMAKOT解析機能を組み込んだ。パスウェイ上で関連するmKIAA抗体のニーズを喚起し共同研究を促進することを目的としている。

【主たる特許】

1. 情報処理装置、情報処理方法及びそのプログラム

特開：2006-185412（2006年7月13日）

出願人：新日鉄ソリューションズ（株）、かずさDNA研究所

2. パスウェイ表示方法、情報処理装置及びパスウェイ表示プログラム

特願：2005-212824（2005年7月22日）

出願人：（株）数理システム、新日鉄ソリューションズ（株）、かずさDNA研究所