マイクロアレイ解析のための自然言語処理による文献検索システム

DNAの配列、遺伝子、蛋白質、化合物などデータベースに蓄積されている情報から、これらの機能に関連した生物医学的な意味を導くためには、これら静的なデータ－ベース情報を統合化し、相互関係を明らかにしなければならない。相互関係を明らかにすることによって、生命がシステムとして機能する相互のメカニズム、あるいは疾患のメカニズムを明らかにすることができ、創薬にもつなげることができる。

近年文献の電子化、データベース化も急速に進んでいることから、我々は、文献情報から必要な情報を信頼度高く抽出し、知識を集約・統合化し、なおかつ分かりやすい形で表示し、しかも予測ができるような自然言語処理技術を応用した生物医学のための文献情報基盤処理システムの構築を目標としている。

その手始めとして、文献から遺伝子、蛋白質等の相互関係を調べ、文献からこれらの相互の関連性を抽出し、相互作用全体のネットワークあるいはパスウェイを構築し、DNAマイクロアレイ解析の結果をシステマティックに解析できるシステムの構築を進めている。具体的には、分散処理環境など様々な計算機環境で、右の特徴を備えたシステムを作成している。

- マイクロアレイ解析に特化したシステムの特徴として、マイクロアレイのプローブ情報からの多数の検索を効率的に行なうための簡単な操作による検索機能
- マイクロアレイのプローブ情報以外にも様々な形式で蛋白質相互作用を検索するための多数のデータを扱うための検索の高速化
- ユーザ指定によって情報を広く収集したり絞り込んだりしながら文献検索マイニングを行なえる機能
- 関係性に関する様々の文献情報をユーザニーズにあわせて検索結果を様々に表示する機能
- 入力された情報と得られたあたらしい情報を様々な研究者のニーズをみたすように様々な形式でネットワークを構築しながら表示する機能
- ネットワークの詳細、およびパスウェイの動的な変化などをユーザが理解しやすいような可視化機能

これらのシステムをオープンラボプロジェクトと連携し、様々なユーザ実際に使用してもらい、創薬の研究開発を効率化・迅速化するためのツール開発を進めている。実験データに応用した成果は、がん学会をはじめ、様々な国際学会、論文などで発表する予定。以上のチャレンジングな技術開発を進め、創薬の研究開発を効率化・迅速化する上で実用に供せる新しいタイプのシステムを作成し、産学連携を進める。

図1 >>生物医学情報の急激な増加

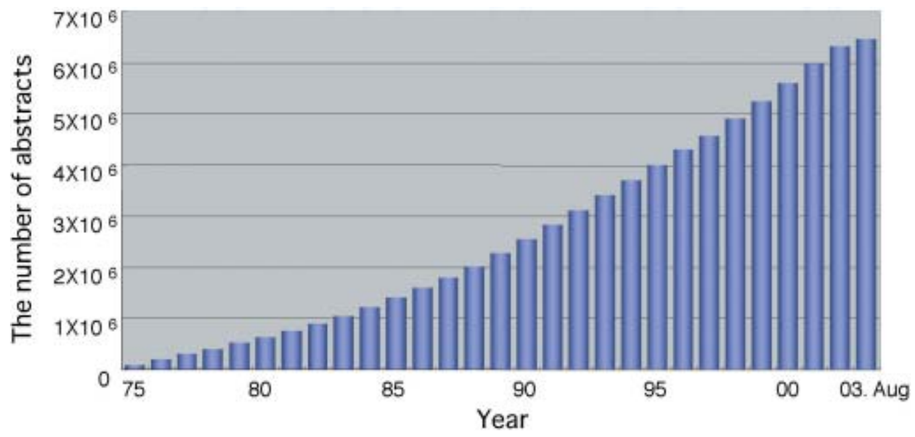
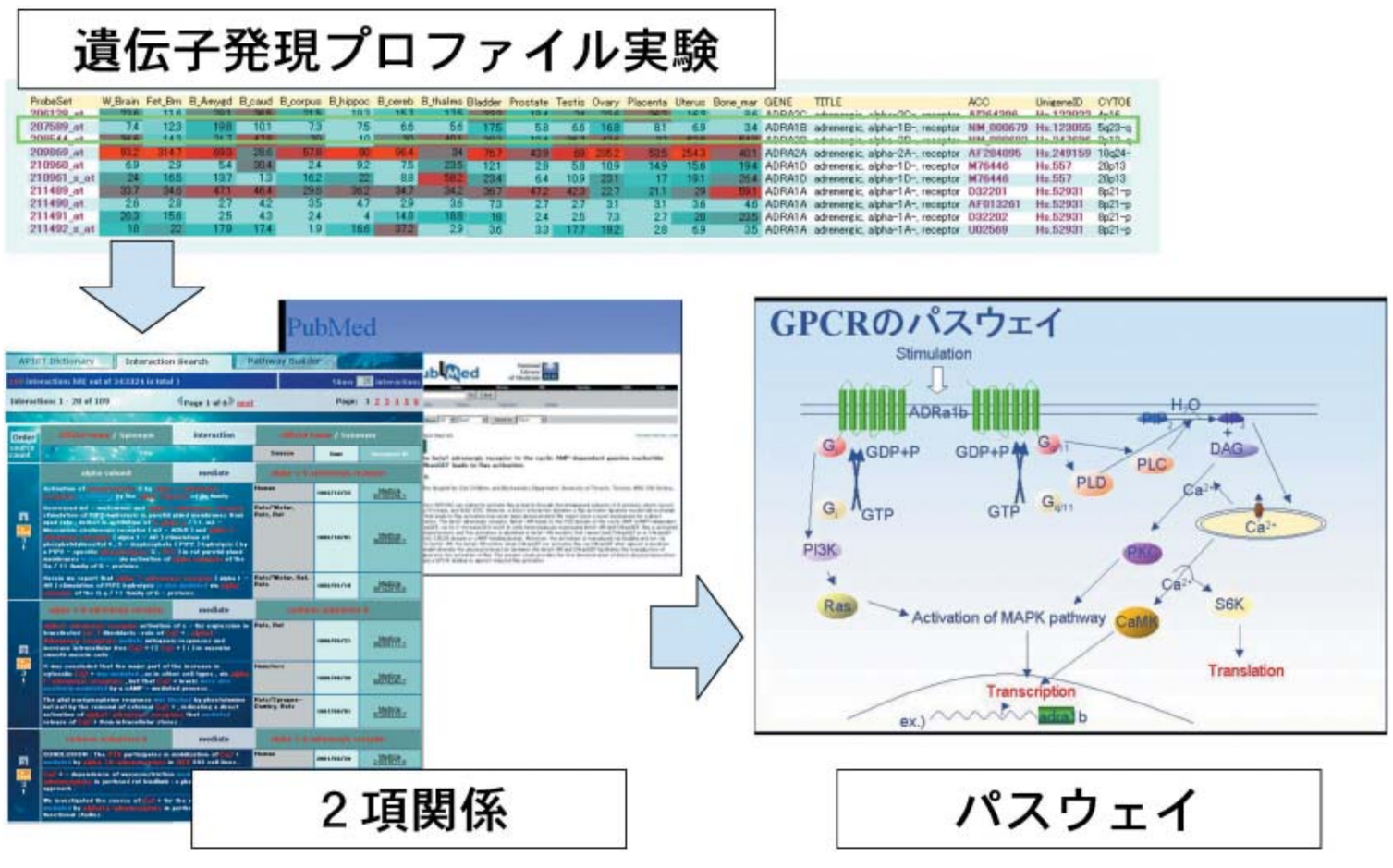


図2 >>遺伝子発現プロファイルデータからのパスウェイ構築



研究者

教授　**油谷 浩幸**

専門分野：ゲノムサイエンス
研究内容：ゲノム科学　機能ゲノミクス　癌のシステム生物学

教授　**井原 茂男**

専門分野：ダイナミカルバイオインフォマティクス
研究内容：文献と実験から知識を統合化して生物をシステムとして理解する、バイオインフォマティクス

井原茂男研究室・油谷浩幸研究／TBIプログラム

図3 >>文献検索システムの概要

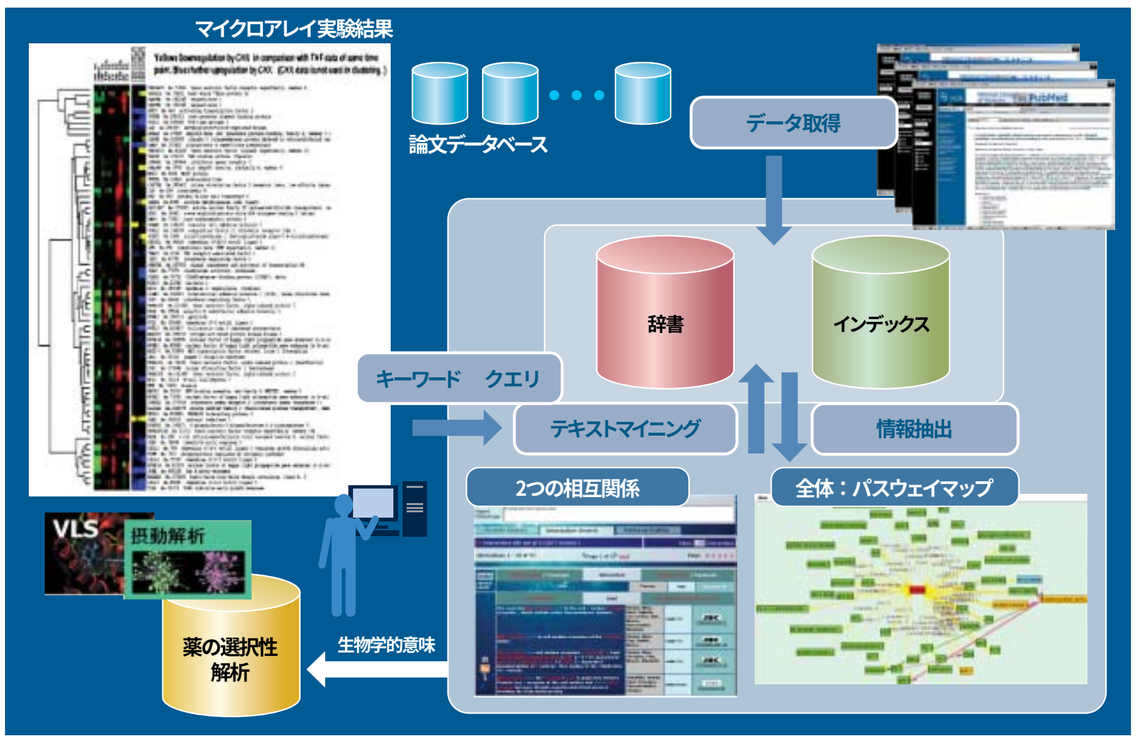
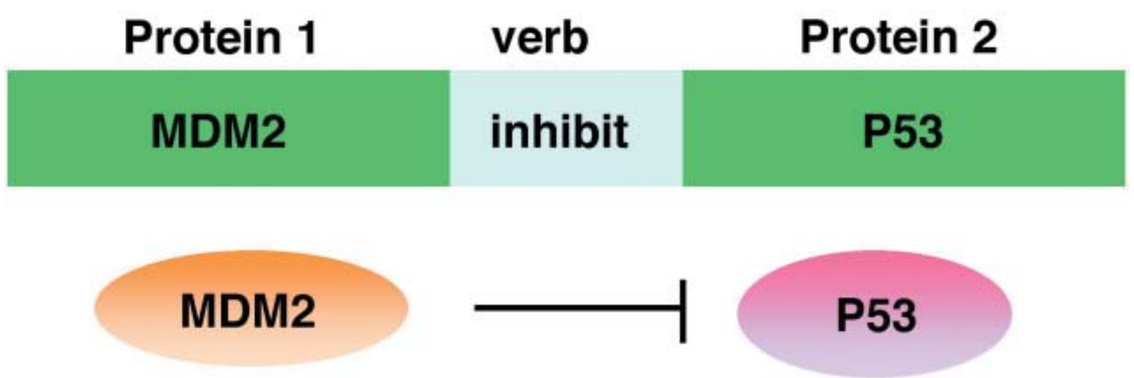


図4 >>テキスト中の「2項関係」の抽出



自然言語処理により蛋白質・遺伝子相互の
（名詞および動詞の）２項関係を抽出

図5 >>クエリーインターフェース

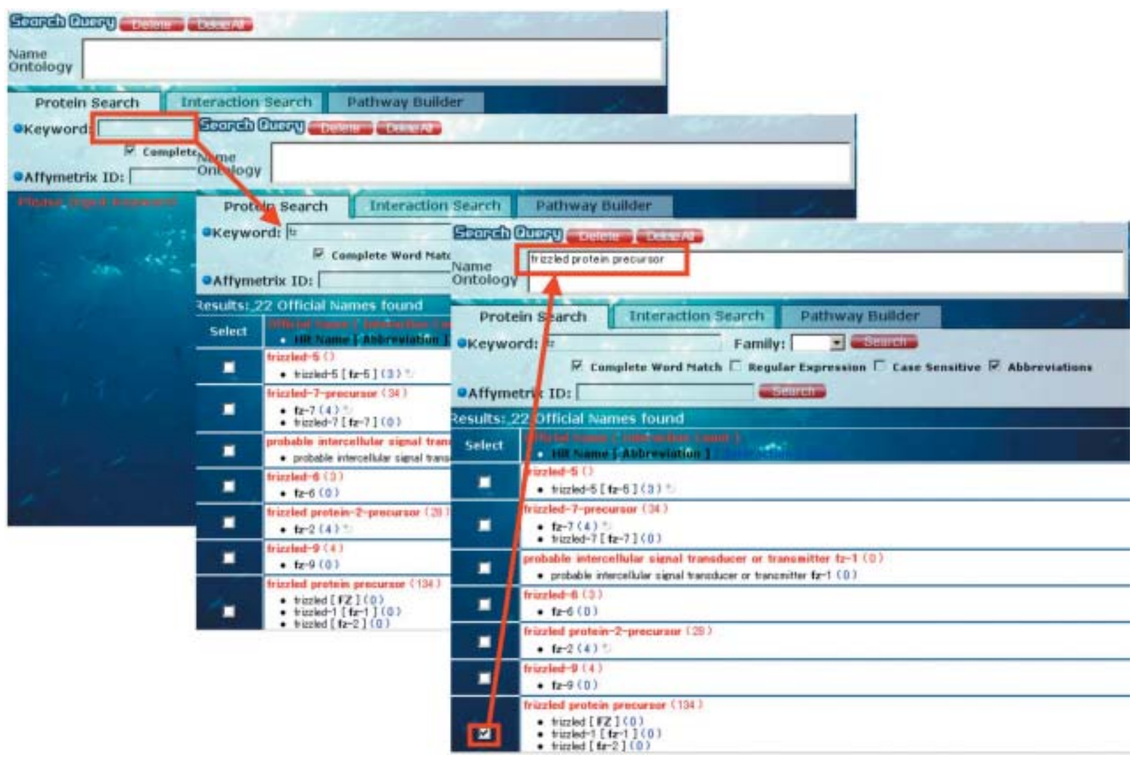


図6 >>パスウェイ構築のインターフェース

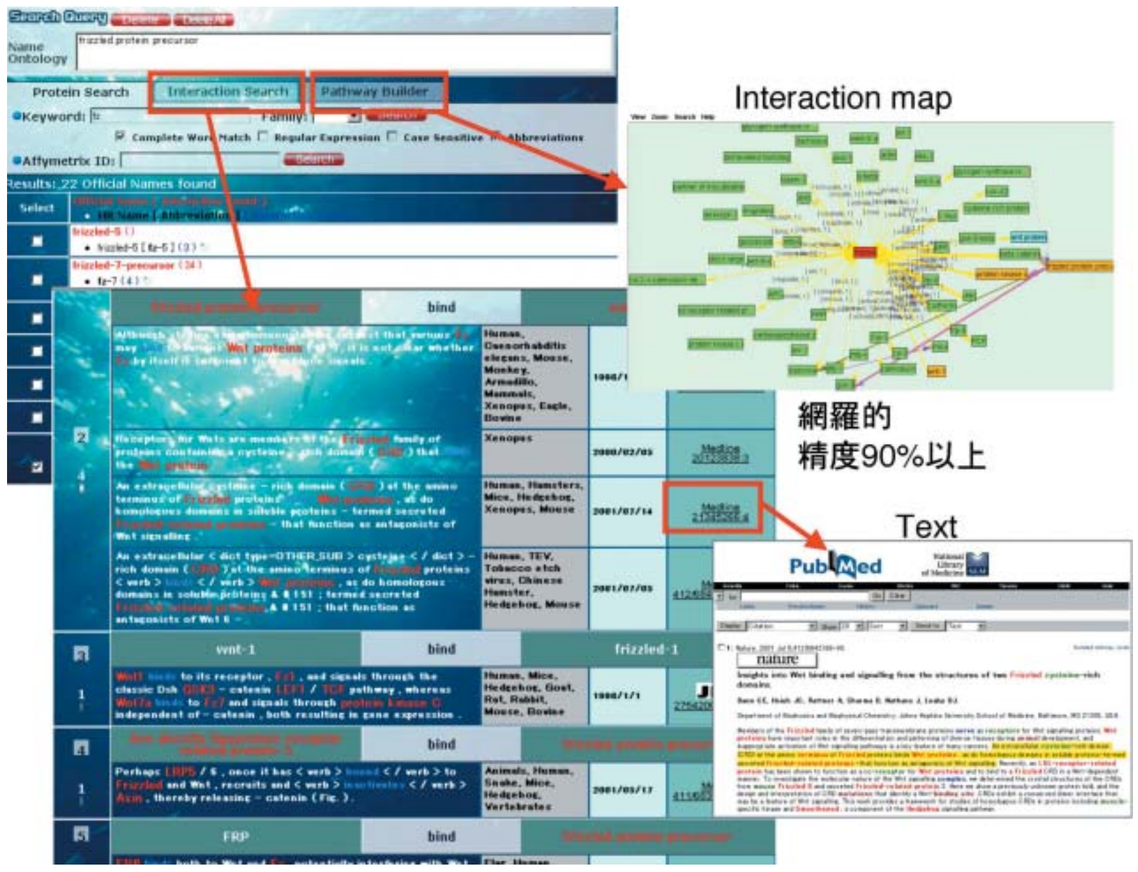


図7 >>ツール活用事例

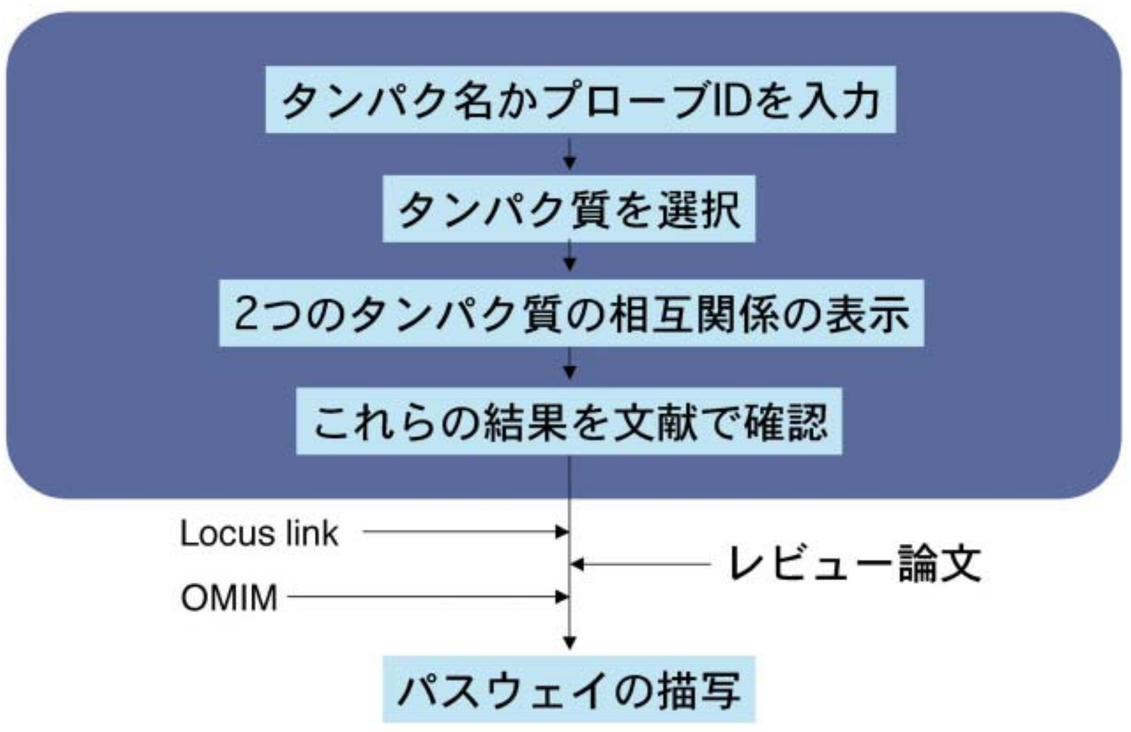
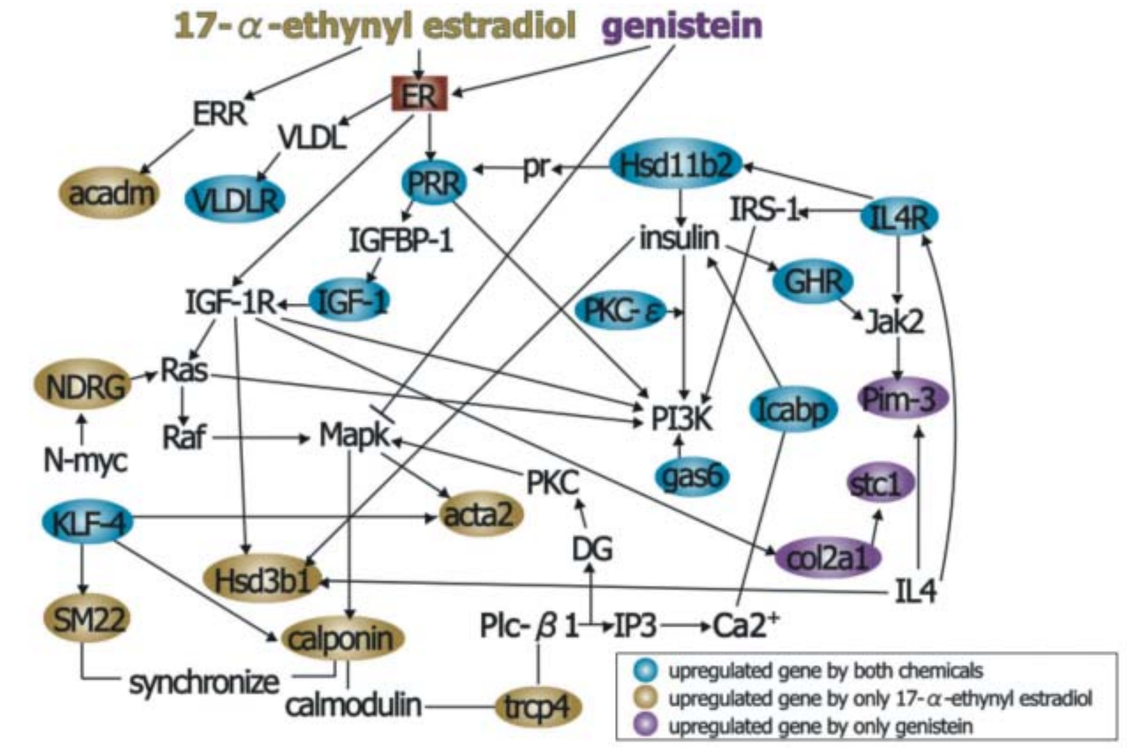


図8 >>薬剤によるパスウェイ変化の予測例



Naciff, J. M. et. al. Toxicology Science 68, 183-199 (2002).