

ライフサイエンス情報統合基盤技術開発

生命情報工学研究センター

【研究成果の概要】

・プラットフォームの利用によりデータベースと解析ツールを組み合わせたシームレスな大規模大量データ解析を可能とする技術開発
 ライフサイエンス研究分野では、その細分化や専門化に伴い、様々なデータベース、解析ツール、予測ソフトウェアが開発され膨大な数に達しており、これらを有効かつ効率よく利用するには様々な課題を克服する必要があります。生命情報工学研究センターでは、これら分散するソフトウェア解析やデータベース検索等を、一つ一つユーザが実行するのではなく、一連の処理の流れを定義することにより、効率的に短時間に実行するためのワークフローに関する技術開発を進め、統合化を目指した実用的な情報基盤開発に取り組んでいます。

【平成22年度計画との関連】

小項目:II-1-(3)-<3> バイオデータベース整備と利用技術の開発
 ・遺伝子や生体分子に関する情報の高度な利用を促進する情報データベースやポータルサイト等を構築する。また、ヒトの遺伝子、RNA、タンパク質、糖鎖情報等の整備及び統合を行うとともに、診断技術等の利用技術を開発する。

【開発技術の用途】

・ライフサイエンス研究分野における専門化・細分化されたデータベースや解析ツールの迅速な変化や新規情報技術へ柔軟に対応できるワークフロー解析

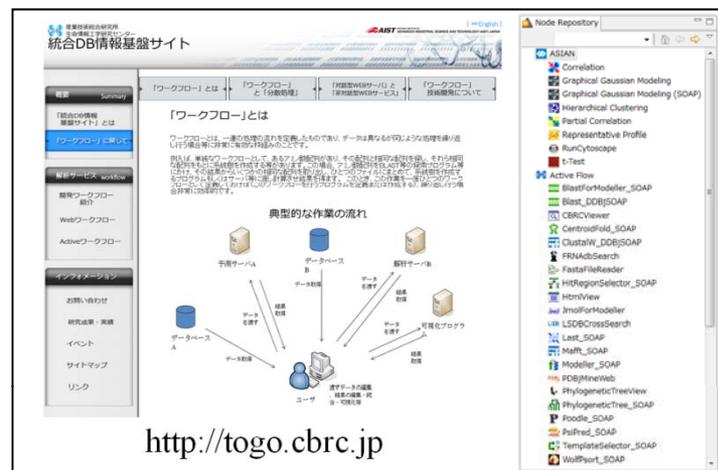


図1. CBRC統合 D B 情報基盤サイト

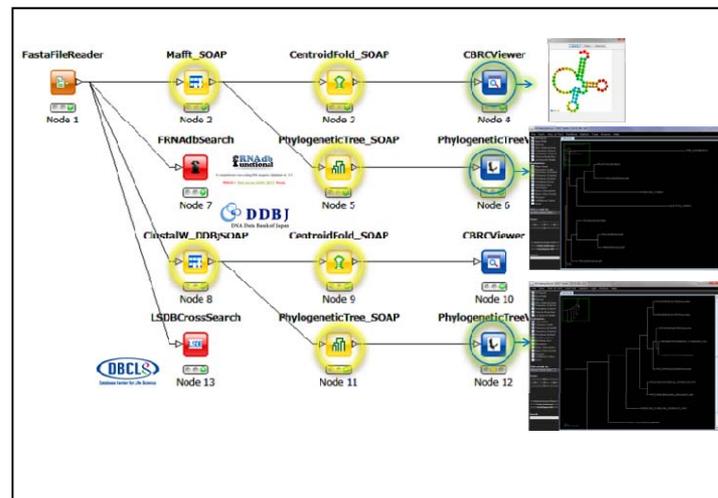


図2 プラットフォームによる解析例

ギガシーケンスデータの高速解析技術の開発

生命情報工学研究センター

【研究成果の概要】

ギガシーケンサーの出力配列を、高速にクラスタリングするアルゴリズムを開発した。本手法は、既存類似手法と比較して格段にメモリ消費量が少なく、特別な計算機がなくとも実行可能である。本手法により、解析に必要な情報量を落とすことなく、効果的に解析対象を絞り込むことが可能となる。得られる生物学的な情報量が飛躍的に増すと期待できる。

【平成22年度目標との関連】

小項目：1-(3)-① 配列情報と分子構造情報を用いた創薬支援技術開発

- 高速シーケンス解析を容易にする基盤整備の一環として、高速シーケンス解析のための重要度の高い基本ツールを提供する。

【開発技術の用途】

- 高速シーケンス解析における基本ツールとして幅広い用途への応用が期待できる：メタゲノム解析、発現解析 (RNA-seq)、エピゲノム解析 (ChIP-seq) など、あらゆる解析に応用可能な重要ツール。

高速シーケンサ



様々な解析

従来法

大規模な
計算機資源が必要



メタゲノム

発現解析

エピゲノム

SLIDESORT

パソコンでも
処理可能



図1 高速シーケンス解析のコストを下げ、より多くの研究室で実施可能とする基盤技術

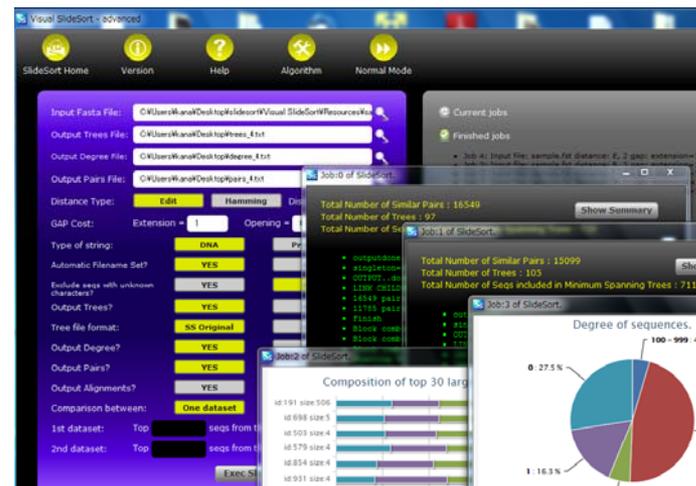


図2 GUIで使いやすさも工夫

高速・高精度ゲノム・アラインメント技術の確立

生命情報工学研究センター

【研究成果の概要】

ゲノム・アラインメントは遺伝子の同定、機能推定と制御メカニズムの解明に不可欠な問題であり、数年前から大規模なゲノム・アラインメントは行われてきた。しかし、従来の技術は繰り返し配列とA+T含量の偏りに弱く、ヒトゲノムの半分ほどの領域や(A+Tに富んだ)マラリア病原菌などに対応できていなかった。そこで、当該センターは従来のデータ構造(固定長文字列のハッシュ表)と異なるsuffix arrayの導入により、繰り返し配列やA+T含量の偏りに頑健なゲノム・アラインメント技術の開発に成功した(Genome Research 2011, Nucleic Acids Research 2010, BMC Bioinformatics 2010 にて発表)。業界標準ソフトウェアと比較した結果、マラリア病原菌のゲノムでは同じ計算精度で100倍以上の高速化が実現できてた(図1)。

【平成22年度計画との関連】

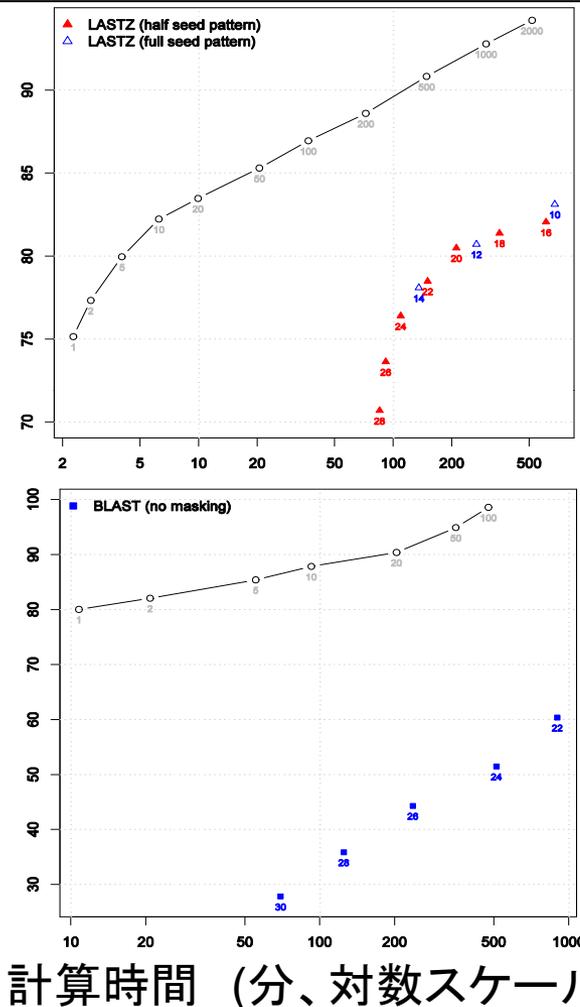
小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

・ヒトの癌細胞や病原菌などの正確なゲノム比較により、遺伝子の機能と発現に重要な領域が分かり、新薬とバイオマーカーの新しい標的の同定に貢献できる。

【開発技術の用途】

・癌細胞におけるゲノム変化と抗癌剤の相性が明らかになり、よりの確な治療と予後予測が期待できる。

精度(最大スコアの回答率)



従来法より約百倍の高速化実現

図1) 開発した手法(O)と競合するBLASTとLASTZの比較。

システム生物学的解析を用いた創薬基盤技術の開発 生命情報工学研究センター

【研究成果の概要】

- 1) 既知ネットワーク構造と計測データとの整合性を統計的に評価する新規技術を開発し、この技術をiPS細胞から得られた遺伝子発現情報に適用することで細胞特異的28制御ネットワークを同定した。
- 2) 代数算法を利用した新規パラメータ推定技術を開発し、シミュレーション実験により従来法に比べ格段に高精度な推定が可能なことを確認した。

【平成22年度目標との関連】

小項目：II-1-(3)-<2> システム生物学的解析を用いた創薬基盤技術の開発

- ・ 既知ネットワーク構造評価法をネットワーク・スクリーニングと命名し、iPS細胞と特徴づける制御ネットワークを同定するとともに、レクチンアレイ計測データの解析も合わせ、細胞特異的な13の糖転移酵素群を同定した。

【開発技術の用途】

- ・ 創薬ターゲット候補推定において、既知の制御ネットワークから特異的に活性化するネットワークを抽出する新しいアプローチを提案
- ・ iPS細胞を含む幹細胞の標準化に向けて、細胞を特徴づけるための合理的な観点を提案。

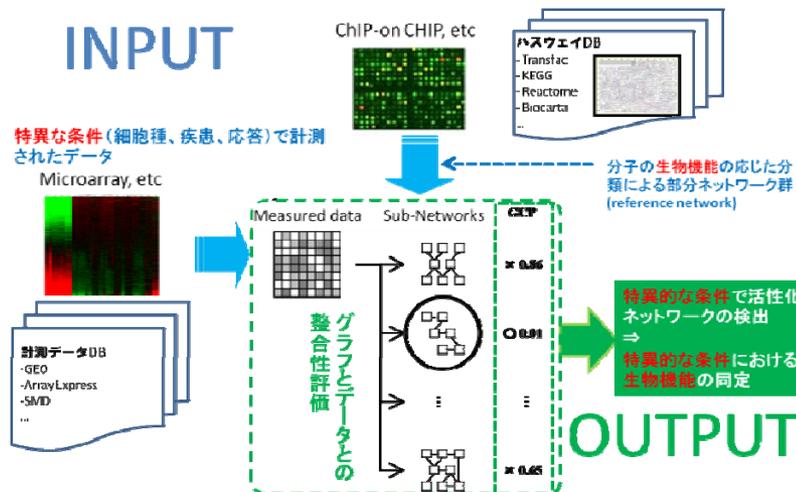


図1. ネットワーク・スクリーニングの概要

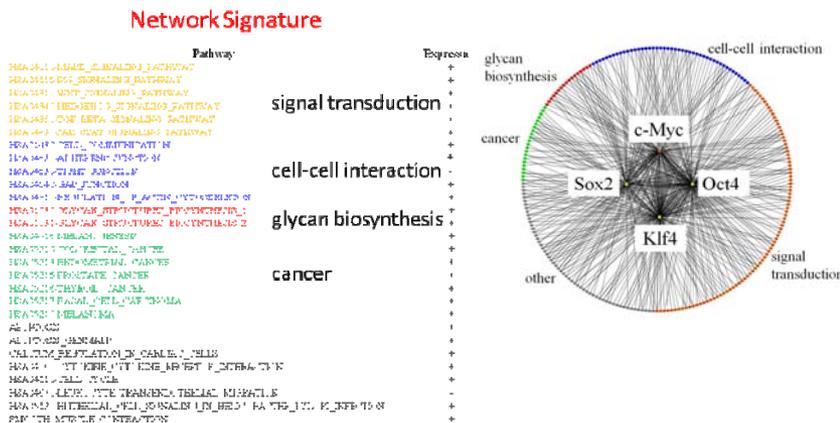


図2. iPS細胞特異的制御ネットワーク

分子動力学シミュレーション技術による化合物可溶化メカニズムの解明 生命情報工学研究センター

【研究成果の概要】

創薬研究において、標的分子への結合活性は良いが、その難溶性のため製剤化に達することのできない問題がある。本研究では、アルギニン分子を混合することで難溶性薬物の可溶化の改善を行うAASS法において、モデル薬物として、没食子酸を例に分子動力学シミュレーションにより、原子レベルでのメカニズム解明に成功した（筑波大との共同研究。J. Phys. Chem. B誌にて発表）。没食子酸が凝集するメカニズム（図1）や、さらにアルギニン分子の効果を等価分子であるリジン分子とメカニズムを比較することで、難溶性薬物改善分子の働きを明らかにした（図2）。

【平成22年度目標との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

- 分子構造情報を用いた分子動力学シミュレーションにより、難溶性薬物の物性予測を行うことで、より品質のより医薬品候補分子の選定に貢献する。

【開発技術の用途】

- 難溶性薬物の可溶化の改善において、難溶性薬物の物性評価を本手法によって事前に予測することにより、メカニズムの理解と製剤化に向けた合理的な物質改善の指針を提案。

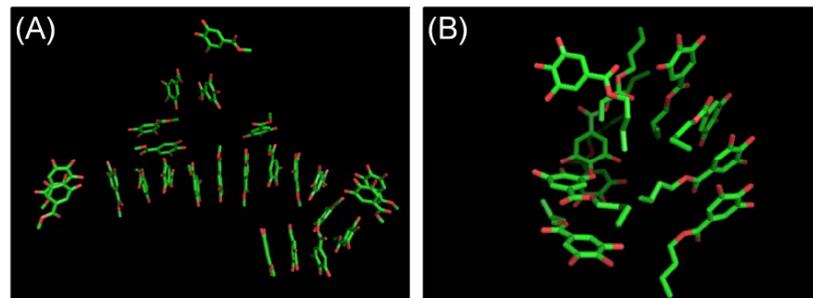


図1．MDシミュレーションによる溶液中における没食子酸構造：メチル没食子酸(A)およびブチル没食子酸(B)

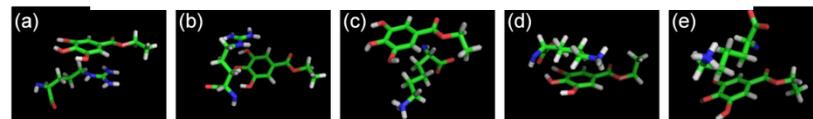
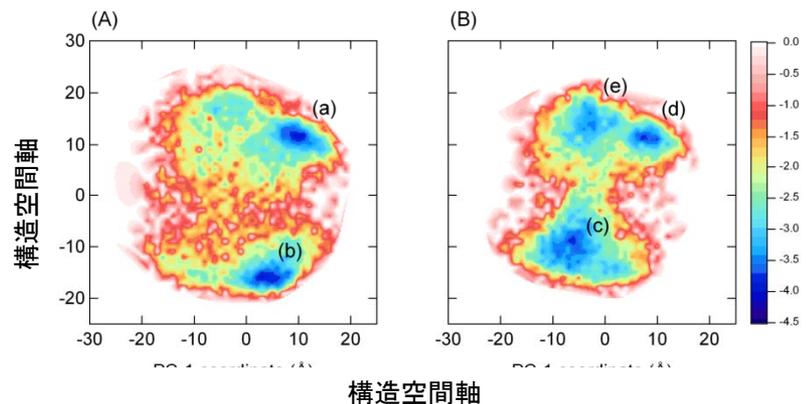
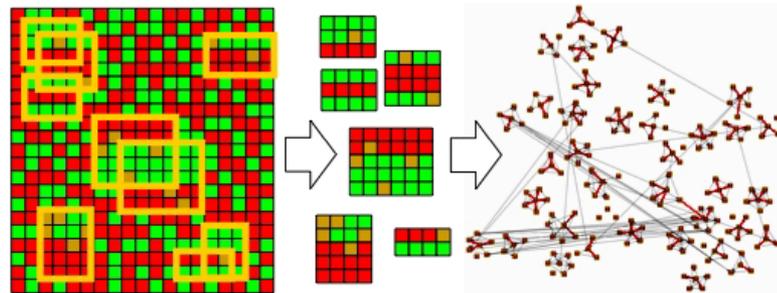


図2．溶液中におけるアルギニン分子 エチル没食子酸会合構造の自由エネルギー空間 (A) と安定会合構造 (a,b)。アルギニン分子に代わってリジン分子と会合した場合を、B, c-eに示す。

高速遺伝子モジュール統合解析システムの最新版SAMURAI2を製薬企業へ販売 生命情報工学研究センター

【研究成果の概要】

2008年に開発・販売開始した数1000件の大量遺伝子発現データから高速かつ網羅的に遺伝子発現モジュール列挙を行う探索プログラムSAMURAI(2008.7.3日刊工業新聞)のバージョンアップ開発を行った。従来に比べ約20倍以上の実行速度の向上の他、時系列データへのウィンドーモジュールサーチ、遺伝子モジュールの有意性の統計評価の他、製薬企業が守秘のため、得られた遺伝子モジュールをパスウェイマップでリアルタイムに表示する機能をインターネットを介さず行うKEGGミニグラフィック機能を搭載した。2010年に実際に大手製薬企業に本プログラムを販売した。



大量遺伝子発現データから従来より約20倍高速に遺伝子モジュールを抽出できるSAMURAI2を開発。遺伝子ネットワーク局所解析へ応用できる。

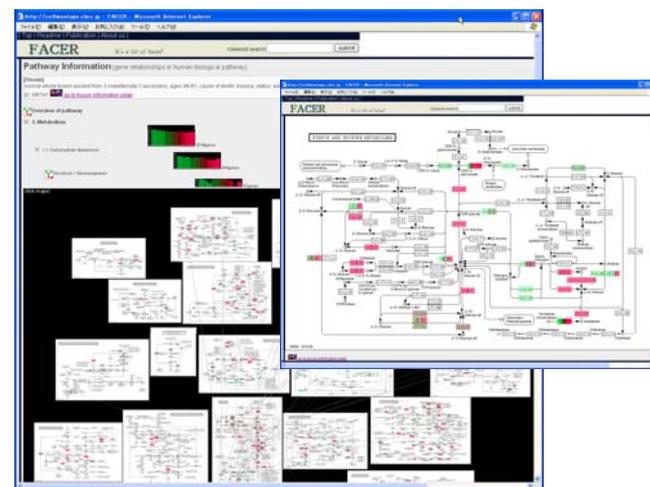
【平成22年度目標との関連】

小項目：1-(3)-② システム生物学的解析を用いた創薬基盤技術の開発

- ・ 転写制御、シグナル伝達、代謝に代表される、細胞内のネットワーク、パスウェイ等の推定やシミュレーションにより、創薬に必要な化合物の設計と合成、標的分子を推定する技術を開発する。

【開発技術の用途】

- ・ 製薬企業で過去に蓄積されていると推定される大量の遺伝子発現データの中からこれまで発見されていないデータマイニングを行うことができ、創薬標的推定やネットワーク局所解析へ応用できる。



製薬企業からの情報漏洩を防ぐため、得られた遺伝子モジュールをインターネットを介さずにパスウェイ解析できるKEGGミニグラフィック機能を搭載した。2010年には実際に本プログラムを大手企業へ販売した。