

サブテーマ名：新規血管構成細胞分化誘導因子を用いた血管再生療法の開発

小テーマ名：データベースの構築とデータ解析法の開発

テーマリーダー：(財)先端医療振興財団、特別研究員、ラルス・マーチン・ヤークト

研究従事者：(財)先端医療振興財団、特別研究員、岡田 光浩

● 研究の概要

DNA マイクロアレイデータ解析には、大量のデータを効率的に処理するプログラムが必要である。しかし、市販のソフトウェアは従来の統計ソフトを改良したものに限られ、そのプログラムの内容は、一般的な統計手法をパッケージングとして用意しているだけであり、生物学的考察を必要とする DNA マイクロアレイデータの解析には適していない。そこで、バイオインフォマティクスの要素を取り入れた DNA マイクロアレイデータ解析プログラムの開発を行った。

1. フェーズ I

(1) 研究の目標

ES 細胞実験モデルを利用した DNA マイクロアレイデータの情報処理・データ解析技術の開発を目標とした。

(2) 実施内容

当研究プロジェクトは、将来の再生医療における基盤技術開発を前提とし、実際の医療現場への血管再生医療用移植材料(細胞)を供給することを目的とした技術基盤の開発を戦略目標と位置付けた。我々はこの細胞運命(分化)決定機序を解析するにあたり、現在最も適した実験材料の1つとしてマウス ES 細胞を実験材料として選択した。

データ解析の実験モデルとしてマウス ES 細胞を用いた *in vitro* 血管細胞分化誘導系を利用し、それらをデータ解析プログラム作成用のサンプルデータ(DNA マイクロアレイデータ)として利用した。プログラム開発用のランゲージは、C++系のQTを採用し、サーバーOSはマルチタスク・マルチユーザー機能で信頼性があり、拡張性の容易なLinuxを採用した。

まず、多種類(多段階)の細胞のサンプルデータ(現在42ポイント)を取得済して、データ解析時にこれらのデータを同時に比較データとして利用した。独立した実験セットのDNAチップデータの比較解析の結果と混合実験のデータを比較解析した。その結果、混合実験データをX軸(並列)に置いた多変量解析手法による解析データのほうが、抽出データの精度が高い結果となった。つまりデータ解析にはアレイデータのData Base化(DB化)が有効であることが判明した。また、データ解析を行う上で有用なマウスゲノム情報(Ensemble)を入手し、データ解析プログラムの中心情報に採用した。この情報は、DNA マイクロアレイ技術の基になった情報であるため、各DNAチップ上のプローブ情報の整合性や適合性を照会でき、解析データの精度評価を可能にした。このゲノム情報の利用は、データ精度評価に利用できるばかりではなく、将来登場すると想定されるDNAチップタイプにもフレキシブルに対応できる(マルチプラットフォーム化が可能)。特にAffymetrix社のGeneChipシステムをplatformに採用したデータ解析プログラムには適している。

さらに、DNA チップデータ間を比較解析するための、ANOVA Sort プログラム（多変量解析手法を利用した）を採用し、信頼精度に基づく自動的並べ替え（ランキング）リスト作成による遺伝子発現情報の自動抽出プログラムを作成した。また、遺伝子情報を利用した生物学的注釈、ID番号、遺伝子ファミリーなどの付帯情報に基づく情報抽出プログラムも作成した。

（3）成 果

DNA マイクロアレイデータ解析用ソフトウェア（データ解析プログラム）の作成

2. フェーズII

（1）研究の目標

マウス ES 細胞由来血管細胞分化系譜の DNA マイクロアレイデータセットをパイロットスタディーとして「DNA マイクロアレイデータを利用した機能的バイオインフォマテイクスシステム」のデータ解析プログラムの開発に成功した。

この研究成果を技術基盤として包括的な再生医学領域の DNA マイクロアレイデータベースをコアとする「ES 細胞分化経路地図」の作成を目標に拡充する。

（2）実施内容

このオリジナルデータ解析ソフトウェアを有効かつ機能的に動かすために、データ解析システムの構造をインターネットを介するクライアント・サーバ型のリレーショナル・データベースシステムを採用し、オリジナル研究ツールとして開発した。

リレーショナル・データベースシステムを採用したことにより、データ解析ソフトウェアは、DNA マイクロアレイデータ解析法を中心とした機能的情報解析システムの端末操作ソフトウェアとしての機能的役割を持ち、複数のエンドユーザーのクエリーに対して必要な対象情報をデータベース（各分子機能関連情報等）から自由に取り出すことが可能となった。

上記 データ解析システムは、次の3つの主要素から構成されている。

1. 複数の解析手法を含むデータ解析ソフト
2. 膨大なデータ量を収める DNA マイクロアレイデータベース
3. 情報処理用サーバ・プロセッサ

上記3点において、多種類の細胞のサンプルデータ（150 ポイント以上）を取得済みであり、データ解析時にこれらのデータを同時に比較データとして利用することで、データ解析精度が実際向上することが判明したため、データ解析には DNA マイクロアレイデータの Data Base 情報を有効活用している。また、我々は Affymetrix 社の GeneChip システムを platform に採用したデータ解析ソフトを開発した。平成15年度には、このデータ処理技術の特許申請を完了した。また、このシステムを利用した血管細胞分化関連遺伝子の *in silico* スクリーニングとマウス胎仔での *in vivo* 発現情報（3次元位置情報）を蓄積している。これまでに血管細胞分化関連遺伝子群の遺伝子発現パターンが数パターン（血管特異的発現パターンも含め）に分かれることを見出しており、血管発生分化に関連した新規候補遺伝子のリストアップもできている。また、このデータ解析システムを利用することで血管細胞の新たな起源を見出せる可能性も推察されている。

また、このデータ解析システムを駆使して、mKIAA 遺伝子群（千葉県 CREATE 共同研究）が血管

細胞分化関連遺伝子群でもあることが見出された。

データベース登録サンプル数の増加に伴い、ユーザー数の増加が生じたため、このシステムは情報安全管理システムの問題上、特定施設内に限定したイントラネットでの使用制限を設けて運用している。さらに、血管細胞以外の ES 細胞由来分化系譜細胞のデータ増加に対応するため、オリジナルデータ解析ソフトウェアには、一般的な解析機能の他に、再生医学研究に適した解析手法も採用した。上記の DNA マイクロアレイデータの蓄積および関連情報のデータベース化は、当研究チームと理化学研究所 発生・再生科学総合研究センター他 外部機関の協力で進めており、それを基にしたデータ解析方法の充実を行っている。

(3) 成果

マウス ES 細胞の試験管内分化誘導系（細胞の分化決定モデル）の構築と分化系譜における細胞の網羅的遺伝子発現解析（遺伝子発現量を定量的に測定する実験系）、外部刺激（分化誘導刺激）または、遺伝子破壊・増幅操作細胞株を用いた全遺伝子発現量変化を測定したデータベースの構築、得られたデータベース情報から分化モデルの全遺伝子発現パターンの自動抽出を行うアルゴリズム、遺伝子発現プロファイルに基づく細胞間の近似性評価法の開発を行った。

また、上記データ解析プログラムを活用して、マウスゲノム情報を利用した網羅的 cDNA および RNA 転写部位決定探索プロジェクト（理研 FANTOM3 プロジェクト）にも参加・協力した。その成果は、新たな RNA ワールドの発見として実を結んだ。(Science. 2005 Sep 2;309(5740):1559-63)

上述の複数のプログラムを作成する理由は、市販の解析ソフトウェアに導入されているデータ解析手法では、通常 一般的な統計学の確率分布（ポアソン分布）を前提に自動的に計算処理されているものが多い。それらによるデータ解析後のデータを用いて実験者は検証を行ってきたが、DNA マイクロアレイデータとウエットデータとの不一致性という問題点が生じ、DNA マイクロアレイ実験の悪評の一要因となっている。我々はデータ解析手法の再検証と最適化を進めてきた。これまでの DNA チップデータのデータベース化による情報活用の結果、「DNA チップを用いた細胞の遺伝子発現プロファイルは 2 峰性分布を示す。」という DNA チップ実験の特性を見出した。この DNA チップ実験の特性を考慮したデータ処理機能（3 種類の統計パラメータを利用）は、データマイニングの際に有用である。特に当プロジェクトにおいては、この DNA チップ特性を利用した統計解析として、遺伝子発現プロファイルに基づいた「ES 細胞分化過程における各細胞間の類似性（近似性）」を評価する機能として応用している。

我々のデータ解析システムは「正確な解析結果を得るには、データ量が多ければ多いほどよい。」という従来の統計学の概念に基づいた網羅的データ解析システムではなく、生物学的特徴を考慮した「大量のデータから正確なデータだけを少量選び出すことが重要。」という概念に基づいている。今後、このデータ解析システムを鋳型としたアフィメトリクス社の DNA チップ（プローブ情報）以外にも対応できるシステム（マルチプラットフォーム化）を作成する。

・オリジナルデータ解析ソフトウェア：

「遺伝子発現プロファイルに基づく細胞間の近似性評価機能」を利用した「2次元 ES 細胞分化経路マップ」描画プログラム。

・特許（基本特許）の出願

3. 今後の展望

5年間に亘る神戸市地域結集型共同研究事業を通じて、我々の研究グループは、プロジェクトの目標とする「再生医療に掛かる基盤技術開発」の土台構築と当グループが創出した技術をコアとした複数の関連技術を活かしたネットワークを基盤技術として神戸ポートアイランド内に整備した。

すでに我々のグループは、DNA マイクロアレイ技術をコアとした独自のインフォマテイクス技術「The eXintegrator system: <http://www.cdb.riken.go.jp/scb/> 参照」を利用し、神戸地域だけではなく、他の複数の国内（関西、関東、他の地方都市圏）主要研究機関や大学をサポートしており、研究プロジェクトの目標通り、我々のグループの研究テーマの特性（蓄積型研究）を活かした「地域での研究・基盤技術創出」にも成功した。上記のように地域創出（神戸）の研究成果物が核となり、周辺研究・技術とのネットワーク形成の仕組み創出することにも成功した。本事業において、我々が想定できる「機能細胞創出に必要な基盤技術」として以下のものを創出・用意した。

1. 機能細胞制御に関する分子レベル（網羅的）の情報データベース
2. 機能細胞作製用 *in silico* セル・マニピュレーション（細胞分化経路検索）プログラム
3. データベースからの抽出情報と各細胞制御分子（候補分子）の機能解析
4. 上記情報の検証と細胞マニピュレーション技術を目的とした細胞レベルの評価系（複数分子のコンビネーション/パラレル評価とハイスループット化）

新規遺伝子の PCT 国際特許申請およびオリジナルデータ解析システム（当研究テーマの成果物）の外部研究機関への導出も完了した。

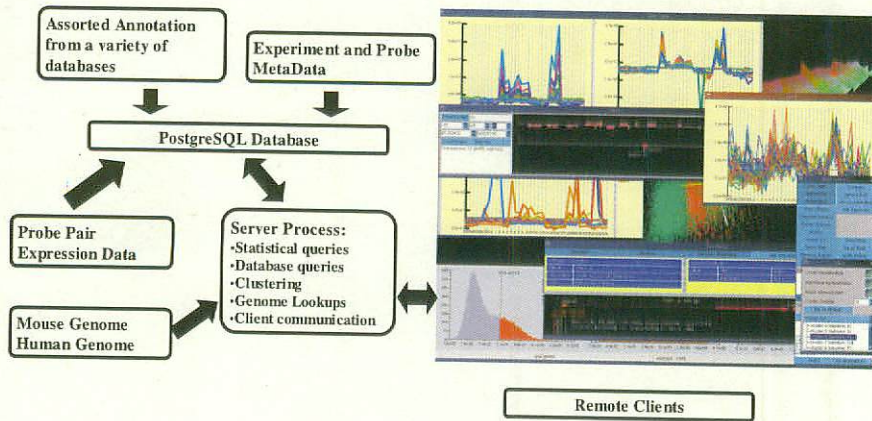
本事業の目標する「研究成果を活かした事業化」という点においては、この研究成果をもって、当研究グループは当初の使命を果たしたといえるだろう。

しかし、「再生医療の実現化」という目標においては、残念ながら、現時点の技術・知識では、活きた細胞を医薬品基準で安全に扱う（操作する）には、まだまだ長期にわたる詳細な研究が必要である。将来 活きた細胞（生体試料）を扱えるような科学的データが蓄積し、我々が細胞を十分に理解することが可能になれば、移植用細胞の人為的操作法の確立が期待できるであろう。生体試料医薬品としての安全性や品質管理基準も同時に確立できるようになると期待できる。

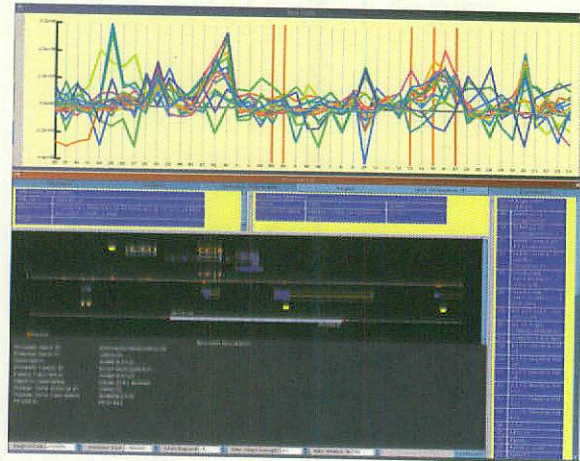
当グループの研究成果は、その実現の可能性と Primary PlatForm を提供するものである。

The eXintegrator System

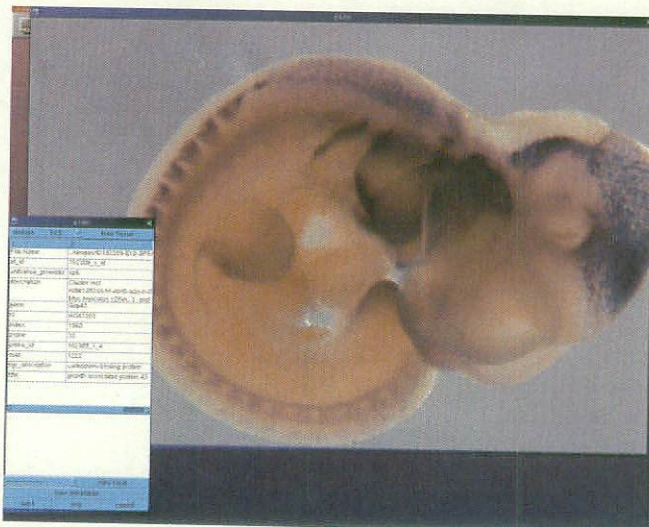
(<http://www.cdb.riken.go.jp/scb/>)



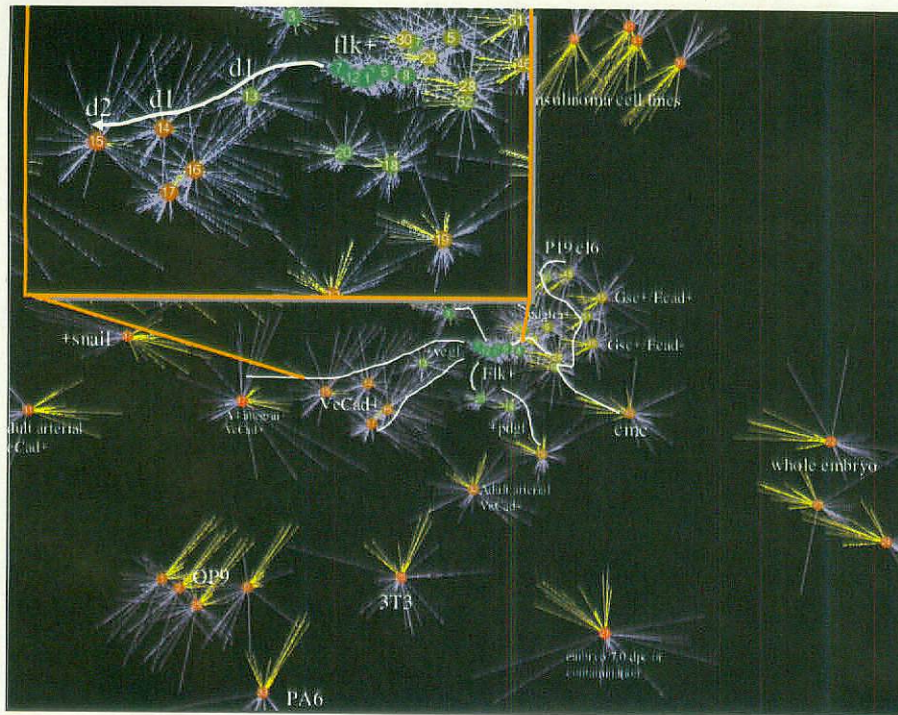
データ解析システムの構造図



オリジナルデータ解析ソフトウェア



Web 実験データのリレーショナルデータベースへの統合



遺伝子発現プロファイル情報を利用した細胞間類似性描画図（2次元マップ）