

研究課題： 生命科学文献からの情報抽出とテキストマイニング技術の開発

研究代表者： 東京大学大学院情報理工学系研究科 宮尾 祐介

研究計画の概要と進捗状況・成果

前年度までの研究において、生命科学文献ベース (MEDLINE) 中の論文抄録の文解析を行い、これをテキストの索引構造に反映した知的情報検索システム (MEDIE、Info-Pubmed) を開発、その成果を国立遺伝学研究所で管理するゲノムネットワークの公式サイトから公開してきた。この2つのシステムは、テキストの意味を文献検索過程に反映することで、従来の文献検索システムの性能 (Precision、Recall) を向上させるものであった。

今年度は、この成果をもとに、より生命科学のニーズに特化したシステムを構築することを目的にして、技術開発を行う。すなわち、(1) 論文抄録から論文全体を取り扱う技術、(2) 単独の生命事象だけでなく、事象のネットワーク (特定のパスウェイ) に関連した情報の抽出技術、(3) 表面上の言語表現に依存しない、より抽象的な生命事象クラスの認識を可能にする技術、を開発し、それらの技術を統合することで、現在の MEDIE の機能を大幅に拡張した文献知識管理システムを開発、公開する。

本年度計画

(1) パスウェイ・データベース構築のためのコーパスの収集：

論文全体を対象としたテキストマイニング技術の開発、また、個別事象から事象ネットワークに関する情報を抽出する技術の開発を系統的に行うための基礎的な言語資料の構築を行う。具体的には、システム・バイオロジー研究機構のSBML研究グループが特定パスウェイのデータベースを構築する際に参照した700件の文献を選び、これらの論文中から、パスウェイ構築に使われた箇所を人手で同定する。

(2) パスウェイ・コーパスへのアノテーションの付与：

(1) で収集された文献中でパスウェイ構築に関連すると同定された箇所に、文の形態品詞、統語木構造、名詞句意味クラスなど、言語処理に必要な情報を付与、および、前年度までの研究で論文抄録を対象に設定した事象クラスの付与、また、論文全体を処理するために必須な共参照関係の付与を行う。論文全体を対象に、これらの豊富なAnnotationを付与した言語資料は、世界的にみても最初のもので、今後のテキストマイニング技術の研究に大きな寄与をするもの

である。

(3) 関係抽出プログラムの開発 :

前年度、GO (Gene Ontology) に基づく生命事象のAnnotationを付与した論文抄録の集合(1000抄録)を構築した。本年度は、この言語資料をもとに文献中から主要な生命現象を記述する箇所を同定する事象の認識プログラムを開発し、表面上の動詞を抽象した、意味的な2項関係で情報単位を抽出する技術を開発する。索引した情報検索システムのプロトタイプを作成する。このプロトタイプは、次年度の研究に引継がれ、次世代公開システムとして整備される。

(4) 意味検索システムMEDIEの高機能化 :

現在、運用中のMEDIEに、(3)の成果を組み込み、生命事象という意味単位の索引構造をもつ検索システムのプロトタイプを作成する。このプロトタイプは、次年度の研究に引継がれ、次世代公開システムとして整備されるものである。

【平成19年度計画予定】

業務項目	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月
パスウェイ・データベース構築のためのコーパスの収集												→
パスウェイ・コーパスへのアノテーションの付与							→					→
関係抽出プログラムの開発									→			→
意味検索システムMEDIEの高機能化												

研究課題：垂直統合型細胞内分子相互作用ネットワーク分析情報基盤の開発

代表研究者： 特定非営利活動法人システム・バイオロジー研究機構 北野 宏明

1. 当初計画・目標

ゲノムネットワークの研究には、ゲノム情報、プロテオーム情報、タンパク質相互作用、遺伝子発現データ、さらには、細胞の動的な挙動に関する理解を行おうとする際には、リガンドアッセイ、リン酸化、一分子観測のデータなども統合的に扱い、秒単位で起きるリン酸化カスケードや細胞内分子移動から、時間単位で起きる発現調節からクロマチン・リモデリングまで記述・表現し、動的な分析をいろいろなスケールで可能にする情報基盤が必要である。本研究計画は、システム・バイオロジー研究機構が開発してきたシステムバイオロジーマークアップ言語 (SBML)、システムバイオロジーグラフィカル表現言語 (SBGN)、CellDesigner 分子相互作用エディターなどを、これらの多様な情報の記述と解析に対応させると共に、垂直統合を実際に行い、その有効性を実証する。システム・バイオロジー研究機構は、酵母や各種哺乳類細胞のシグナル伝達系のネットワークを記述すること、および、ゲノムネットワークを基盤とした生物学研究の為の情報基盤の構築と実証を行う。

2. 現状、進捗状況、成果

計画はおおむね順調に推移しており、SBML Level-2 Version 2 の最終仕様の決定、SBGN Level-1 の Draft 仕様の決定に近づきつつある。特に、SBGN は、新たな標準化作業であるにも関わらず、大きな関心と多くの参加を得て進められている。同時に、これらの標準仕様案を実際に実装したソフトウェア、CellDesigner の開発も進めているが、これは 19 年度に第 4 版をリリースできる状態になった。このソフトウェアは、タンパク間相互作用から遺伝子制御関係までを統合的に、しかも、相互作用のメカニズムまでも詳細に記述することができるほぼ唯一のソフトウェアである。本年 4 月に発行されたシステムバイオロジー関連のソフトウェアに関する調査 (Klipp, et al., Nature Biotechnology) で、CellDesigner はもっとも評価の高いアプリケーションであるという結果が出ている。実際、ダウンロード数は 2 万件に到達しようとしており、国際的に多くのユーザを抱えるソフトウェアに成長している。

CellDesigner の 3.5 リリースでは、ゲノムネットワークのデータベースへのアクセスなどが強化されたが、4.0 では、ゲノムネットワークプロジェクトで同定される遺伝子間の制御関係の記述の強化を行うと共に、Plug-In などを可能とする機能強化を行う。これらの研究開発の結果、現在、もっとも機能的に優れ、国際的に「定番」と考えられるソフトウェアとそれに実装される国際標準が形成されつつあり、これは、ゲノムネットワークプロジェクトの大きな成果であると考えられる。

3. 本年度計画、当初計画・目標に対して達成の見込み

本年度計画：

本年度には、(1) SBGN Level-1 の仕様定義の決定・公開、(2) 動的ネットワーク記述と解析能力強化への柔軟性を向上させる Plug-In 機能を導入した CellDesigner 4.0 の開発・リリース、(3) 次世代版 CellDesigner の基本設計への着手、(4) SBML の強化、(5) CellDesigner アプレ

研究課題： 新技術を基盤とした革新的遺伝子解析システムの開発

研究代表者： 東京工業大学大学院生命理工学研究科 関根 光雄

1. 当初計画・目標、途中計画変更等あればその旨記載

計画の変更は特にありません。

2. 現状、進捗状況、成果

すでに本研究グループは、塩基部無保護法とケミカルライゲーションを用いる感度と精度が極めて高い新しい SNP 検出法を開発した。さらに、これまでミスマッチ塩基対形成により、間違っ
て検出してしまふ遺伝子診断の根源的問題点を解決するために、保護 DNA プローブを開発して
きた。これによって、正確にマッチ塩基対のみ形成できるプローブ分子の合成ができるようになった。
今年度は、総合的にこれらの新技術を実用化する段階にあると言える。

3. 本年度計画、当初計画・目標に対して達成の見込み

塩基部無保護 DNA 合成法に基づく迅速・高精度の新規ゲノム解析技術の開発では、ケミカル
ライゲーションを活用する新しい SNPs 検出法を開拓したことにより、ほぼ目的を達成することが
できた。したがって、今年度では、若干の検出条件の改良をしたあと、研究成果を論文としてまと
める方針である。縦軸の研究グループの中に利用者がいれば、積極的に共同研究を展開したい。フェ
ライト含有磁気ビーズ担体を用いる迅速遺伝子検出法の開発では、新型磁気ビーズに化学合成した
2種類の保護 DNA プローブを導入することを検討する予定である。保護 DNA プローブの合成ユ
ニットがほぼ完成したことにより、この研究も年度内に実施できるものとおもわれる。超高精度塩
基識別能人工塩基をもつ DNA チップの開発では、基本技術となる保護 DNA プローブ合成で、キ
ャップ化反応中のグアニン塩基部位の副反応の問題を解決し、保護 DNA プローブ法の優位性を示
す。

mRNA を標的にした RNA チップの開発では、基盤技術である 2'-O-シアノエチル化された RNA
の合成法をさらに改良し、合成ユニットを大量に得る手法を確立する。

4. プロジェクトの中での研究の位置づけ

本研究グループは、縦軸の研究者に対して、新しい遺伝子検出技術を提供し、共同研究を推進す
べき立場にあると考えている。そのため、縦軸の研究グループに積極的に情報提供し、利用者がい
れば、進んで共同研究を展開したい。

【平成19年度計画予定】

区分	4	5	6	7	8	9	10	11	12	1	2	3
①塩基部無保護DNA合成法に基づく迅速・高精度の新規ゲノム解析技術の開発	←→											
②フェライト含有磁気ビーズ担体を用いる迅速遺伝子検出法の開発	←→											
③超高精度塩基識別能人工塩基をもつDNAチップの開発	←→											
④ mRNAを標的にしたRNAチップの開発	←→											
⑤新技術の評価	←→											

研究課題： ショットガン戦略による高分解能メチル化ボディマッピング

研究代表者： 東京大学 大学院新領域創成科学研究科 伊藤 隆司

1. 当初計画・目標

DNAメチル化情報を単一塩基レベルの解像度を保ったままでゲノムワイドに収集するバイサルファイブ・ショットガン・シーケンシング法(BSS)を確立し、ゲノムネットワークの理解に欠かせない高分解能メチル化ボディマップ作成の技術基盤を確立する。

2. 現状、進捗状況、成果

①BSS法の確立

454シーケンサーGS20によるBSS法のフィージビリティを昨年度のパイロット実験で実証。明らかになった問題点を踏まえて、制限酵素消化ゲノムDNAに直接454用アダプタを付加した上でbisulfite処理を行ない、その産物を試験管内転写の利用で均一かつ効率的に増幅する鋳型調製法を確立。

②Neurosporaをモデルとしたwhole genome BSS

上記手法によるNeurosporaのBSSを進行中。GS20によるデータ収集を間もなく終了の予定。

③BSSデータ高速解析法の開発研究

弱いホモロジーをシードに、bisulfite変換に合致するミスマッチを許容しながらアラインメントをダイナミックに伸展する新アルゴリズムを考案、それに基づくプログラム開発に着手。

④BSSデータベースおよび表示システムの開発

BSS法によるメチル化データベースのスキーマと表示システムの設計が終了、開発に着手。

⑤BSSデータとHM-PCRデータの比較統合

BSSデータのリファレンスとしての転写因子遺伝子プロモータのHM-PCRデータを蓄積中。

3. 本年度計画、当初計画・目標に対して達成の見込み

①BSS法の基本は確立済みなので、今後は実データ生産フェーズ(前半はNeurospora、後半はヒト)で更に洗練する予定。また微量鋳型の増幅法も検討。

②Neurosporaのwhole genome BSSは、リード長の長い454シーケンサーの新機種FLXによるデータも加えて、予定通りに年度内に達成の見込み。その後はヒトサンプルに着手する予定。

③BSSデータ高速解析法は年度前半にプロトタイプを開発し、上記Neurosporaデータの解析で実用レベルにまで洗練する予定。

④BSSデータベースおよび表示システムは、昨年度的设计案に基づく開発を年度内に達成の予定。

⑤BSSデータとHM-PCRデータの比較統合の為のHM-PCRボディマップ作成は予定通りに継続。

4. プロジェクトの中での研究の位置づけ

転写制御を考える際に無視できないエピジェネティック情報をゲノムワイドに収集する先端技術を開発・提供することを通して、より高度なゲノムネットワークの描出に貢献する。

【平成19年度計画予定】

区分	4	5	6	7	8	9	10	11	12	1	2	3
BSS法の確立												
1) ヒト鋳型調製用の洗練							←					→
2) 微量鋳型からの増幅法検討	←											→
Neurospora whole genome BSS												
1) データ収集	←						→					
2) データ解析と取り纏め				←								→
BSSデータ高速解析法の開発												
1) 基本システムの構築	←						→					
2) 実用レベルへの洗練				←								→
BSSデータベースおよび表示システムの開発	←											→
BSSデータとHM-PCRデータの比較統合	←											→

研究課題：精子幹細胞の遺伝子改変によるがん疾患モデルラットの作成

研究代表者：京都大学 大学院医学研究科 篠原 隆司

1. 当初計画・目標

ラット精子幹細胞での遺伝子改変を実現し、がん疾患研究のためのモデルラットを作成する技術の実現を目指し、ほ乳類一般における個体レベルでの遺伝子改変を可能とする技術の開発することを目的とする。

2. 現状、進捗状況、成果

ラット精子幹細胞を免疫不全マウスを仮親にすることでラット子孫の作成を行うことに成功した (Proc. Natl. Acad. Sci. USA 103, 13624-13628 (2006))。これによりラットを飼育するスペースの削減、子孫作成のスピードの短縮につながることを期待される。この成果は異種の精巣内でも精子形成が正常に進行し、そこから生じた精子を利用し正常な個体発生に利用できる点を示した点で画期的なものであった(ラットとマウスの遺伝的な距離はヒトとチンパンジーとほぼ同程度と考えられています)。

3. 本年度計画、当初計画・目標に対して達成の見込み

本年度は遺伝子導入および培養条件の改善、長期培養の影響の解析、薬剤選択技術の開発の3つの技術を統合することを試みる。

1) 遺伝子導入および培養条件の改善:

これまでにラット精子幹細胞クローンを取ることで自体についてはある程度成功しているが、まだ数多くのクローンを簡単に短期間で取るための条件設定を行う必要がある。

2) 長期培養の影響の解析

ラット精子幹細胞がどの程度の期間長期培養に耐えることが可能であるのかを確認し、どのくらいの期間であれば精子形成能力を保持することができるのかを調べる。

3) 薬剤選択技術の開発

これまでの精子幹細胞における薬剤選択のプロモーターにはマウスの場合にはpgkプロモーターを使ってきた。しかしながら、ラットでは薬剤選択におけるneomycinに対する感受性が非常に高く、マウスと同濃度で行うとすべての細胞が簡単に死滅してしまうという問題がある。また精子幹細胞内ではウイルス系のプロモーターは抑制されるという問題点があるために、現段階で薬剤選択に最も適したプロモーターを再検討する必要がある。

4. プロジェクトの中での研究の位置づけ

昨年度は子孫の作成についての成功を報告したが、本年度はラット精子幹細胞の遺伝子改変の

基本になる、培養条件の改善、長期培養の影響の解析、薬剤選択技術の開発の3つを柱として研究を進める。この実験結果が予定通り得られれば、次のステップとしてノックアウトベクターの遺伝子導入、スクリーニングへのステップへと研究を進めていく予定である。

【平成19年度計画予定】

区分	4	5	6	7	8	9	10	11	12	1	2	3
①遺伝子導入および培養条件の改善		←										→
②長期培養の影響の解析	←											→
③薬剤選択技術の開発		←										→

研究課題： 転写因子に対する抗体の遺伝子免疫による迅速作製システムの開発

研究代表者： 東京理科大学基礎工学部 千葉 丈

1. 当初計画・目標、途中計画変更等あればその旨記載

本申請研究では、従来の免疫法の問題を解決する方法として、申請者の開発してきた遺伝子免疫法の転写因子抗体作製への最適化を行い、ChIPアッセイなどの転写因子の機能解析に有用な転写因子に対する抗体を迅速に作製するシステムの開発を目指す。さらにその過程で、抗体が必要であるにもかかわらず抗体が作製されていないか、抗体があってもChIPアッセイなどができない重要な転写因子を3年間で20～30種類選んで、それらに対する抗体を作製することを目的とする。計画変更はない。

2. 現状、進捗状況、成果

初年度（平成18年度）は下記のように、目標をほぼ達成することができた。

- 1) 迅速な抗体作製法に最適化した遺伝子免疫法が、転写因子に対する抗体作製にも応用できることを、RXRG および RORB 遺伝子を用いて証明した。
- 2) 抗体のない新たな3種類の転写因子に対する抗体作製を上記1)の免疫スケジュールで行い、3種類全てに対する抗血清を調製できた。また、この中の1種類の抗体はChIPに利用できることを証明した。

さらに、当初の目標を上回る以下の成果を得ることができた。

- 3) 調製された抗ヒトRXRGおよびRORB抗体は、マウスのRXRGおよびRORBとも強く結合したので、本研究で作製された抗体を用いて、ヒトとマウスの同じ転写因子の機能解析を併行して行うことが可能になった。

3. 本年度計画、当初計画・目標に対して達成の見込み

- 1) 転写因子に対する抗体の迅速作製システムの開発
我々が確立してきた遺伝子免疫法をさらに最適化し、ChIPアッセイなどに有用な転写因子抗体の迅速作製システムの改良を進める。
- 2) 転写因子に対する抗体の作製
転写因子の研究グループとの密接な共同研究のもとで、THP-1細胞で発現する重要な転写因子を新たに10種類程度選択し、それらに対する抗体を作製する。すでにいくつかの抗体の作製に着手している。

4. プロジェクトの中での研究の位置づけ

THP-1細胞で発現する重要な転写因子に対する有用な抗体を作製できるので、GNPの発現クラスターワークショップに貢献できる。また、横浜市大・鈴木正則先生、理化学研究所・鈴木治和先生との連携を深める。