

## 5. プロジェクトの中での研究の位置づけ

ゲノムネットワークプラットフォームにおけるデータベースからの発見的情報利用技術の開発を、大学共同利用機関法人情報・システム研究機構国立遺伝学研究所、国立大学法人東京大学、特定非営利活動法人システム・バイオロジー研究機構と共同で実施する。この目的のため、長浜バイオ大学では、選択的スプライシングによる蛋白質相互作用の変化を推定するアルゴリズムに関わる研究開発を実施している。

### 【平成20年度計画予定】

4月	}	マウスの AS データとの関連づけと進化的保存性のある AS の抽出 発現データとの対応づけとそれに基づく重要度の高 AS の抽出
5月		
6月		
7月	}	ヒトおよびマウスの AS データの更新
8月		
9月		
10月	}	ユーザーのデータから AS を解析する Web サーバー機能の実装 ANNALS へ重要度の高い AS の提示
11月		
12月		
1月	}	
2月		
		3月

## 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

### 1 当初計画・目標

東京大学での研究は、文献情報を有効活用する技術を開発・提供することで、生命科学研究における発見の情報利用に資することを目的とする。特に、意味に基づく言語処理技術を使うことにより、従来のテキスト・マイニング技術の限界を超えた性能のシステムを開発する。具体的には、Pathway データベース構築のための基礎データ作成、および Pathway 作成補助システムの構築、意味検索システム MEDIE と Info-Pubmed の恒常的な運用システムの構築、テキストマイニングシステムモジュラリティの向上を目指して、データ作成及びシステム構築を行う。

### 2 現状、進捗状況、成果

- 2.1 **Pathway データベース構築のためのコーパス収集とアノテーション付与**: 前年度からの継続として、抄録からフル論文を対象を拡大し、これに対して意味アノテーション (Term Annotation と Event Annotation) を付与する作業を開始した。
- 2.2 **Pathway作成補助システムの構築**: 1.1の結果を用いて、関連文献の収集、情報抽出、Pathwayの視覚化機能を持ったPathwayの構築支援システムを構築中である。
- 2.3 **意味検索システムMEDIE、Info-Pubmedの恒常的な運用システムの構築**: プロジェクトの最終年度として、これまで試行的にサービスしてきた文献検索システムを恒常的に運用するためのシステムの整備を行う。とくにあらたな文献の追加を半自動的に行うシステムを構築し、近く実用化する。
- 2.4 **テキストマイニングシステムモジュラリティの向上**: これまで作成してきたテキストマイニングシステムをコンポーネント化し、さまざまな組み合わせで使えるツールキットとして整備し、幅広いユーザに提供する。
- 2.5 **GENIA コーパス**: GENIA コーパスとして統語情報・意味情報が既にタグ付けされている Medline 抄録文約 2,000 件に対して、専門用語の意味クラスを拡張すると共に、事象アノテーションのスキーマを改訂して、データの修正作業を行った。

### 3 成果予定

- 3.1 **システムの公開**: 現在公開中のシステム (MEDIE、Info-Pubmed) に定期的な新テキストの追加などのメンテナンス機能を付加することにより、恒常的なサービスシステムとして運用していく。また、ユーザ教育など、広報活動を充実させ、より多くの研究者に使われるサービスシステムとする。
- 3.2 **ソフトウェア・ツール**: プログラマブル・アクセス機能を持ったソフトウェアのインターオペラビリティを確保するための国際コンソーシアムを構築し、われわれの仕様を国際標準として定着させる。
- 3.3 **論文発表**: ツール、コーパス、言語処理技術に関して、Bioinformatics、ACL などの論文誌・国際学会に発表する (5 編)。また、個別生命科学研究グループとの共同研究の成果を生命科学系の論文誌に発表する (2 - 3 編)。

#### 4 本年度計画、当初計画・目標に対して達成の見込み

Pathway データベース構築のための基礎データ作成及び Pathway 作成補助システムの構築は、平成 20 年 12 月末を目途に完成を目指し、公開する。MEDIE、Info-Pubmed の恒常的運用システムの構築は、平成 20 年秋を目途に完了を目指し、プラットフォーム内でのシステム運用への実用化を図る。また、テキストマイニングシステムのコンポーネント化及び、ツールキットの整備については、完了したのちから随時公開し、ユーザに提供する。

#### 5 プロジェクト内での研究の位置づけ

本分担課題は、基礎生物学から医・薬学等の応用分野に至るまで、統合的な情報の提供を可能にするデータベースからの発見的情報利用技術の開発を目的とし、共同研究環境としてのヒトゲノムネットワークプラットフォームを実現して、ヒト生体分子ネットワーク情報の収集・解析と、利用者への公開を実現するために、パイプライン技法を中心とした高速なデータの連携・解析・可視化手法を実現するものである。

#### 【平成20年度計画予定】

区分	4	5	6	7	8	9	10	11	12	1	2	3	
①パスウェイ・データベースの構築のためのコーパスの収集とアノテーション	←—————→												
②パスウェイ作成補助システムの構築	←—————→												
③意味検索システムMEDIE, InfoPubmedの恒常的運用システムの構築	←—————→												
④テキストマイニングシステムモジュラリティの向上	←—————→												

## 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、これは本 20 年度に第 4 版をリリース予定である。このソフトウェアは、タンパク質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。昨年 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 万件を超え、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った。これらを中心とした強化版を CellDesigner 4.x として開発・リリースする。既に、4.0 β 版で多くの機能を取り入れてリリースしており、正式版として、ベータ版でのユーザからのフィードバックを吸収する。

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

本年度には、(1) SBGN Level-1 の仕様定義の決定・公開、(2) SBGN Level-2 の仕様定義、(3) 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.x の開発・リリース、(4) 次世代版 CellDesigner の基本アーキテクチャの設計、(5) SBML の強化、(6) CellDesigner アプレットの開発が、主な項目となる。これらの研究開発に加え、さらなる普及とユーザのニーズ把握のために、ユーザ会議や開発者会議を行うことを検討する。

本研究は詳細な分子間相互作用を記述し、データベースへのアクセスやシミュレーションを可能とする統合ソフトウェア CellDesigner へ集約し実装している。また、ビジュアル表現の標準として SBGN を提案し、標準化作業も進んでいる。このプロジェクトの実証として、Toll-like receptor

と RTK のシグナルカスケードの大規模 MAP を完成させた。また、CellDesigner 自体のユーザー数も、世界的に 2 万ダウンロードを超え、標準ソフトとしての地位を築きつつある。ここでの最大の狙いは、本分野での国際標準に関する主導権の維持であり、定番ソフトウェアの開発によって、今後の研究のリーダーシップを握ることである。これを達成することは十分可能であると考えられる。

本年度計画予定を、別紙スケジュール表に示す。

#### 4. 成果予定、論文発表、特許出願、データ産出状況等

秋頃 SBGN 関連論文を発表すべく、準備を進めている。また、8 月にスウェーデンで開催の ICSB 2008 において、関連イベントとして SBGN ワークショップ（ユーザ会議等）を実施の予定である。

#### 5. プロジェクトの中での研究の位置づけ

本計画は、ゲノムネットワークプロジェクトの情報インフラストラクチャーの中で、ネットワークの記述の部分を担当するものである。特に、プロジェクトの後半で遺伝子間の制御関係のネットワークの同定が進み、その記述が必要になってくる場合に、本課題での成果を利用してネットワーク記述を行うことになる。これらの情報を国際標準の形式で発信することは、成果普及の観点から、非常に重要な役割を担っていると考える。

特に今年度の開発で転写制御関係の記述方法の基本方針が固まったことで、GNP で解明されたネットワークを表記する方法に大きな前進があった。実際に、この表記方法に基づいたソフトウェアのテスト版を理研チームに提供しており、今後、GNP でのネットワーク表記により重要な役割を果たすと期待される。

#### 【平成20年度計画予定】

区分	4	5	6	7	8	9	10	11	12	1	2	3
(1) SBGN Level-1仕様定義・公開	←————→											
(2) SBGN Level-2仕様定義					←————→							
(3) CellDesigner 4.xの開発・リリース	← Ver 4.0 ———→					← Ver 4.x ———→						
(4) 次世代版CellDesignerの基本アーキテクチャの設計	←————→											
(5) SBMLの強化	←————→											
(6) CellDesignerアプレットの開発	←————→											

## 次世代ゲノム解析技術の開発

**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

新しい DNA の合成技術を駆使して高精度かつ迅速遺伝子解析法の開発をおこない、ゲノムネットワークプロジェクトにおけるゲノム解析の技術的な発展を図る。

### 2. 現状、進捗状況、成果

本研究グループでは、新技術として塩基部無保護法、保護 DNA プローブ法を活用して、精度の高い遺伝子検出法を開拓してきた。とくに、天然塩基の塩基識別能を凌駕する”保護型”人工塩基を組込んだ DNA プローブ分子の開発も4種類創成することに成功した。現在、この人工塩基の合成ユニットのコスト軽減を図り、実用化に向けた研究を展開している。現在、この保護プローブを導入した RNA チップの創成にも研究を進展させている。

### 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。保護 DNA プローブを実用化について、現在、合成ルートの見直しや、合成中間体の結晶化など、大量合成に必要な最終的な技術の確立を行っている。人工塩基としてグアニン塩基のアミノ基をカルバモイル化したものが優れた性質をもっていることを明らかにしてきたが、唯一溶解性に問題が残されている。そこで、この問題を解決することを最終年度の目標としている。そのために、0-6 位にシリル系の保護基を導入した合成ユニットを構築し、この改良した合成ユニットを用いて4種類の任意な塩基配列をもつ保護 DNA プローブの創成と応用を行う予定である。すでに、シリル系の保護基の目処もついていることから、目標を達成できる予定である。RNA チップに関しては、年度末までに完成させる予定である。

### 4. 成果予定、論文発表、特許出願、データ産出状況等

改良された合成ユニットを用いる保護 DNA プローブ法については間もなく成果があがると思われる。また、2-チオウラシルや 4-*N*-カルバモイルグアニンなどの個々の保護型人工塩基の塩基識別能と塩基対形成能に関する論文は9報すでに発表しているが、関連論文も逐次発表の予定である。特許に関しては、保護 DNA プローブとその合成に必要なリンカーに関するものを含めてすでに4件申請した。

## 5. 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20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ト・ショットガン・シークエンス法(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 法基礎条件の改良

二量体形成が最も少ない方法として非リン酸化アダプターを利用する方法を採用し、バイサルファイト変換後の微量鋳型を試験管内転写によって均一に増幅する方法を確立した。また、アダプター内にバイサルファイト変換効率をモニターするチェッカー配列の導入を行った。

#### ②Neurospora をモデルとした Whole Genome BSS

上記で開発した方法を、モデル系として取り上げた Neurospora ゲノムに適用し、3種類の制限酵素 (MboI, FatI, Tsp509I) を用いた鋳型調製を行なった。それらを GS-20 および FLX による配列決定に供した結果、約 98 万の BSS リードを得た。

#### ③BSS データ高速解析法の開発研究

前年度の開発研究によって明らかになった配列データマッピングの問題点を検討し、新しいマッピング・アルゴリズムの開発を進めた。その結果、BSS生データ (Query配列) にC-to-T変換を行ない、それをForward鎖とReverse鎖のそれぞれにC-to-T変換を行なったゲノム配列に対してBLASTを行ない、アラインメント後にC-to-T変換を復元してメチル化パターンを判定する手法を開発した。

この手法を、上記(1)で得た98万リードのデータに適用したところ、用いる酵素および機種によっても異なるが平均92%の効率でマッピングに成功した。その結果、Neurosporaゲノム全体の48.8%にあたる19Mbを平均5.7の重複度でカバーするBSSデータの産生に成功した。

### 3. 本年度計画、当初計画・目標に対して達成の見込み

- ①制限酵素を用いない鋳型調製法の開発。年度前半に達成の見込み。
- ②454以外のシーケンサーの導入。年度内に達成の見込み。
- ③ヒトゲノム特定領域のBSSとヒトゲノム用マッピングのチューニング。年度内に達成の見込み。
- ④BSSデータベースおよび表示システムの開発。年度内に達成の見込み。
- ⑤BSSデータとHM-PCRデータの比較統合。年度内に達成の見込み。



**4. 成果予定、論文発表、特許出願、データ産出状況等**

Neurospora のBSS解析は、現在、論文を準備中である。制限酵素を用いない調製法の進行状況によっては、データを追加して発表の予定である。

ヒトに関しては年内に 21 番染色体上の特定領域に関する BSS 解析を予定。

特許に関しては、制限酵素を用いない鑄型調製法の開発過程で汎用性のある要素実験技術を開発したので、出願の検討を開始した。

**5. 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20年度計画予定】**

区分	4	5	6	7	8	9	10	11	12	1	2	3
制限酵素を用いない鑄型調製法の開発	←-----→											
454以外のシーケンサー導入	←-----→											
ヒトゲノム特定領域のBSSとヒトゲノム用マッピングのチューニング	←-----→											
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

昨年度の結果を利用し、ラット精子幹細胞へレンチウイルスを導入し、免疫不全マウスの精巣内へ移植することでマウスを仮親にすることでトランスジェニックラット子孫の作成を行うことに成功した（投稿準備中）。これにより遺伝子改変した精子幹細胞から遺伝子改変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が可能になった。

ラット精子幹細胞の試験管内での変異導入については薬剤選択法が確立し、現在2つの遺伝子トラップラインの個体化に成功した。また、相同組み替えによる目的遺伝子の破壊についてはこれまでサザンブロット、PCRにより解析したものの、まだ相同組み替え体を得るには至っていない。

### 3. 本年度計画、当初計画・目標に対して達成の見込み

本年度は遺伝子トラップ個体のホモ化、顕微受精の効率化と相同組み替え体のGS細胞の樹立、その個体化に取り組む。

#### 1) 遺伝子トラップ個体のホモ化

これまでにマウスの異種移植法を用いて、2つのラインでラット精子幹細胞クローンからの個体化に成功している。ホモ個体の作成のために、SPF環境へラットを移動して、現在交配中。

#### 2) F1-GS細胞の樹立、顕微授精の効率化

より効率よくラット産仔を得るために、SDラットと純系ラットとのF1系統からのGS細胞の樹立、顕微授精における卵子の系統差を評価する。

#### 3) 相同組み替え体のGS細胞の樹立

昨年までに確立したGS細胞の遺伝子選択法を適用し、引き続きノックアウト細胞の樹立のためにHPRT、オクルジン、p53遺伝子ノックアウトベクターの導入、薬剤選択、PCR/サザンブロット法による遺伝子解析を行う。組み替え体が得られ次第、異種移植法を利用して早期に個体化を行う。