

「機能性RNAプロジェクト」

事後評価分科会資料

資料 5-1

健康安心イノベーションプログラム

「機能性RNAプロジェクト」

事業原簿

【公開版】

作成者

新エネルギー・産業技術総合開発機構  
バイオテクノロジー・医療技術開発部

— 目次 —

概要	.....	i ~ v
プロジェクト用語集	.....	viii ~ x
I. 事業の位置付け・必要性について	.....	1
1. NEDO の関与の必要性・制度への適合性	.....	1
1.1 NEDO が関与することの意義	.....	1
1.2 費用対効果	.....	2
2. 事業の背景・目的・位置づけ	.....	2
II. 研究開発マネジメントについて	.....	3
1. 事業の研究開発目標	.....	3
1.1 研究開発目標	.....	3
1.2 目標設定の理由	.....	3
2. 事業の研究開発計画	.....	3
2.1 研究開発項目①		
「機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発」	.....	3
2.2 研究開発項目②		
「機能性 RNA 解析のための支援技術・ツールの開発」	.....	4
2.3 研究開発項目③		
「機能性 RNA の機能解析」	.....	4
3. 研究開発実施の事業体制	.....	5
3.1 事業体制	.....	5
3.2 研究開発の運営管理	.....	6
4. 研究開発成果の実用化に向けたマネジメント	.....	10
4.1 知財戦略	.....	10
4.2 成果実用化を推進するマネジメントの取り組み	.....	10
5. 情勢変化への対応	.....	11
5.1 研究テーマおよび研究体制の見直しと強化・発展	.....	11
5.2 中間評価の受審と対応	.....	11
5.2-1 中間評価	.....	11
5.2-2 中間評価結果	.....	12
5.2-3 対応を要する指摘事項への対応	.....	13
III. 研究開発成果について	.....	1.1
1. 事業全体の成果	.....	1.1
①機能性 RNA の探索解析のためのバイオインフォマティクス 技術の開発	.....	1.2
②機能性 RNA 解析のための支援技術・ツール開発	.....	1.4
③機能性 RNA の機能解析	.....	1.6

2. 研究開発項目毎の成果	
2. 1 機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発	..... 2.1.1
2. 1. 1 機能性 RNA に特化したバイオインフォマティクス技術の開発	..... 2.1. 5
2. 1. 2 ゲノム配列からの機能性 RNA の網羅的予測	..... 2.1.41
2. 1. 3 機能性 RNA データベースの構築	..... 2.1.53
2. 2 機能性RNA解析のための支援技術・ツールの開発	..... 2.2. 1
2. 2. 1 RNA マススペクトロメトリー法の開発	..... 2.2. 3
2. 2. 2 機能性 RNA の高感度検出システムの開発	..... 2.2.14
2. 2. 3 RNAの新規合成基盤技術開発と化学分子設計	..... 2.2.41
2. 3 機能性RNAの機能解析	..... 2.3.1
2. 3. 1ヒト疾患に関連する機能性 RNA の迅速で高効率な同定	..... 2.3.5
2. 3. 1. 1 ヒト疾患に関連する機能性RNAの同定とその機能解析	..... 2.3.5
2. 3. 1. 2 機能性RNA解析に基づくゲノム医学研究	..... 2.3.40
2. 3. 2 機能性 RNA に関する基盤的知見の獲得と それを基にした機能性RNA同定	..... 2.3.70
2. 3. 2. 1 機能性 RNA 候補の多面的選別法の確立と機能解明	..... 2.3.70
2. 3. 2. 2 マイクロRNAの作用機構の解明	..... 2.3.94
2. 3. 2. 3 アンチセンスRNAの機能解析	..... 2.3.104
IV. 実用化の見通しについて	..... 15

(添付資料)

1. イノベーションプログラム基本計画
2. プロジェクト基本計画
3. 技術戦略マップ(分野別ロードマップ)
4. 事前評価関連資料
5. 成果発表・論文・特許出願リスト

# 概要

作成日 平成22年5月11日

プログラム名	健康安心イノベーションプログラム						
プロジェクト名	機能性 RNA プロジェクト	PJコード	P06011				
担当推進部/担当者	バイオテクノロジー・医療技術開発部/主査 伊豆本 義隆						
0. 事業の概要	本プロジェクトでは、バイオインフォマティクス(BI)の活用による機能性 RNA を推定する技術の開発、機能性 RNA 解析のための支援技術・ツールの開発、及び機能性 RNA の機能を解析することにより、本研究分野における我が国の優位性の確立を目指す。また、本プロジェクトにおける成果を素早く特許化することにより、将来的な医療・診断分野における新産業の創出に貢献する。						
I. 事業の位置付け・必要性について	生体における機能性RNAの機能を解析し、その機能の抑制や促進による生体応答を総合的に把握する基盤研究をさらに強力に進めることは、再生医療や疾患治療等の実現に資する重要な課題であり、「健康安心プログラム」の一環として本プロジェクトを実施する。このように基礎的・基盤的研究であるので、NEDOの関与が必要である。						
II. 研究開発マネジメントについて							
事業の目標	機能性RNA候補をゲノム上から推測するバイオインフォマティクス技術を確立する。また、機能性RNAを高感度、定量的かつ網羅的に捉える新しい方法、手法の確立と、機能性RNAをゲノムワイドに解析するためのツールを確立する。これらのツールを活用して、機能性RNAの機能解明に取り組む。また、ヒト疾患に関連する機能性RNA及び発生・分化などをはじめ細胞機能に重要な働きを示す数十個の機能性RNA候補の機能解析を行い、医薬品開発や再生医療等に有用な基盤知見の取得や、基盤技術の構築を目指す。						
事業の計画内容	主な実施事項	H17fy	H18fy	H19fy	H20fy	H21fy	
	BI活用による機能性RNAの推定	←				→	
	支援技術・ツール開発	←				→	
	機能性RNAの機能解析	←				→	
	調査研究	←				→	
成果とりまとめ			↔				
開発予算 (単位:百万円)	会計・勘定	H17fy	H18fy	H19fy	H20fy	H21fy	総額
	一般会計	605	899	798	778	759	3,839
	特別会計	-	0	0	0	0	0
	加速予算	-	0	51	0	0	51
	総予算額	605	899	849	778	759	3,890
開発体制	経産省担当原課	経済産業省 産業技術環境局 研究開発課 経済産業省 製造産業局 生物化学産業課					
	プロジェクトリーダー	プロジェクトリーダー (独)産業技術総合研究所 バイオメディシナル情報研究センター 研究技術統括 渡辺公綱 サブリーダー 東京大学大学院新領域創成科学研究科(産総研情報生命科学センター長 兼務) 教授 浅井潔 東京大学大学院工学系研究科 教授 鈴木勉 (独)産業技術総合研究所 バイオメディシナル情報研究センター チーム長 廣瀬 哲郎					
	委託先 (参加企業)	(社)バイオ産業情報化コンソーシアム (株)三菱総合研究所 みずほ情報総研(株) 株式会社インテックスシステム研究所 (株)DNAチップ研究所 (株)ノバスジーン (株)島津製作所 オリパス(株) ヤマサ醤油(株) 日本新薬(株) ジェノダイブファーマ(株) 東レ(株) 協和発酵キリン(株) 大塚製薬(株) 日立ソフトウェアエンジニアリング(株)					
	共同研究実施先	北海道大学 弘前大学 山形大学 東京大学 慶應義塾大学 東京工業大学 千葉大学 東海大学 北陸先端科学技術大学院大学 京都大学 大阪大学 武庫川女子大学 岡山大学 徳島大学 (独)産業技術総合研究所 (独)理化学研究所					
情勢変化への対応	研究テーマおよび研究体制の見直しと強化・発展 ○「研究進捗状況ヒアリング」【H19(2007)年1月】における見直し						

	<p>研究開始から約1年半が経過した、平成19年1月23日と24日の2日間にわたり、全分室・共同実施先に対して「研究進捗状況ヒアリング」を行い、実施体制の見直しを行った。テーマ「尋常性乾癬に関わる RNA の機能解析」については、当初予定していた臨床検体の入手ができておらず、今後の入手の目処も立たないため、平成18年度をもって本テーマは中止することとした。この結果、ジェノダイブファーマ株式会社（分室10）と東海大学（共同実施先）を研究実施体制から外した。また、テーマ「機能性 RNA を in vivo で計測するシステムの開発」については、標識した RNA を細胞内に取り込ませる技術の確立に手間取り、研究が大幅に遅れてしまった。当面は、本テーマは中断することとし、計画通り進捗しているテーマ「機能性 RNA の高感度検出システムの開発」の方に研究資源を集中することとした。その他のテーマについても、評価結果を平成19年度の予算配分に反映させた。</p> <p>○加速予算の投入【H19(2007)年4月】</p> <p>一方、テーマ②-1「RNA のマスペクトロメトリー（質量分析技術）の開発」は、本プロジェクトを支える重要なツール開発であり、世界的に見ても日本発のオリジナルな技術である。実施計画をはるかに上回るペースで開発が進んでいることから、NEDO の「加速財源制度」を活用し、平成19年4月にフーリエ変換型質量分析装置を導入した。本装置は、2005年にサーモエレクトロン社が発売したもので、質量分解能が飛躍的に高い。本プロジェクトが築き上げた優位性を確保するために、いち早く本装置を導入することとした。</p> <p>○中間評価以後の見直し【H21(2009)年4月】</p> <p>テーマ①-2「ゲノム配列からの機能性 RNA の網羅的予測」における「比較ゲノムによるゲノム相同領域の抽出」(三菱総研担当)については一定成果が得られ、抽出した領域の解析に研究開発の重点が移ったため、プロジェクト途中で終了した。</p>
<p>中間評価結果への対応</p>	<p><b>【結果抜粋】</b></p> <p>順調に研究を進めている。興味深い結果も創出しており、後半は更なる成果が期待される。</p> <p><b>【主な改善点と対応】</b></p> <p>①バイオインフォマティクス技術については、開発した予測技術の有効性を明確にすべき。</p> <p>(対応)機能解析グループとも連携しつつ、検証実験を計画。</p> <p>②支援技術・ツールの開発においては、従来技術と対比した上での活用法、残された解析課題などを明確に。</p> <p>(対応)開発技術の必要性の明確化を行い、テーマを絞込み、予算を集中化。</p> <p>③機能性RNAの機能解明については、良い芽を出している研究を選択し、実用化を目指した研究を行うべき。</p> <p>(対応)産業競争力に寄与しうる先進的な基盤技術になる テーマ、競合優位性を獲得しうるテーマに絞り込んで予算を集中化。</p>
<p>Ⅲ. 研究開発成果について</p>	<p>本プロジェクトでは、以下の3つの研究開発項目について推進し、それぞれ目覚ましい成果を導いた。概要は以下の通りである。</p> <p><b>研究開発項目①「機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発」</b></p> <p>○新規機能性 RNA 発見に必要な、二次構造を考慮した RNA 配列の比較・整列について数々の新規手法を開発し、精度、速度で世界最高性能を達成した。</p> <p>○機能性 RNA の機能予測に必要な、配列群の共通二次構造予測を高精度に行う手法と、二次構造モチーフを配列群から抽出するグラフ理論を用いた新規アルゴリズムを開発した。</p> <p>○ゲノム配列中からマイクロ RNA を従来法よりも精度良く抽出する手法を開発し、330個の新規マイクロ RNA 候補を発見した。</p> <p>○ヒトゲノム中のシンテニー領域、ヒトゲノム同士の比較から得られる相同領域等から、二次構造を考慮した配列解析技術の適用によって 20 万箇所以上の機能性 RNA 候補を抽出し、有望な 1 万領域以上についてマイクロアレイ発現解析を開始した。</p> <p>○新規機能性 RNA 発見のプラットフォームとして、独自の情報解析結果も含めた機能性 RNA データベースを開発した。</p> <p><b>研究開発項目②「機能性 RNA 解析のための支援技術・ツールの開発」</b></p> <p>○RNA マスペクトロメトリーの開発に成功し、サブフェムトモルの RNA 分子を直接解</p>

	<p>析することに成功した。プロジェクト最終目標値を大幅に上回る成果である。この手法を用いて、piRNA の末端修飾構造や miRNA の選択的安定化機構を明らかとした。また、miRNA の直接プロファイリングに成功し、診断技術としての実用化に大きく前進した。また、核酸医薬の薬物動態が可能になり、今後は世界標準を志向した基盤技術としての確立を目指す。</p> <p>○RNA マスフィンガープリント法の開発に成功し、RNA マススペクトロメリーで得られた RNA 断片の質量情報のみを用いて、ヒトおよびマウスゲノムの表裏(6 ギガ塩基)から RNA の遺伝子領域を迅速に特定することに成功した。</p> <p>○RNA の微量解析を支援するツールとして、RNA 自動精製装置(往復循環クロマトグラフィー)の開発に成功した。</p> <p>○RNA 医薬品の開発を目的とした新規化学合成技術(特許出願済)の開発に成功した。この手法により、長鎖 RNA(110 塩基)の合成に成功し、生物活性を示すことができた。高品質かつ安価な RNA を供給するシステムが整いつつある。</p> <p>○微量なマイクロ RNA の発現解析を高感度で行うため、Photo-DEAN 法および、MPEX 法の開発を行い、数アトモルのマイクロ RNA を定量的に解析することに成功した。</p> <p><b>研究開発項目③「機能性 RNA の機能解析」</b></p> <p>○ ヒト間葉系幹細胞、マスト細胞から 1500 種以上の新規 miRNA を同定し特許出願した。幹細胞分化やマスト細胞の脱顆粒、癌細胞増殖抑制に関わる生理活性を有する miRNA とその標的 mRNA を 10 個以上発見した。miRNA のノックアウトマウスを作製し、miRNA が脳下垂体機能である排卵に関与することを初めて示した。</p> <p>○ iPS 細胞の作成効率を著しく上昇させる miRNA を同定した。</p> <p>○H-InvDB から約 100 種の組織特異的または核内局在 ncRNA を同定した。その中から免疫抗原提示やボディーパターンニング、肝細胞癌抑制能、間葉系幹細胞分化に関係する機能性 RNA 候補を同定した。核内 ncRNA 解析のための新手法を開発し、細胞構造構築や細胞周期制御を行う ncRNA 機能を発見した。10 種類の疾患タンパク質と ncRNA の相互作用を明らかにした。</p> <p>○ショウジョウバエの生殖細胞特異的な piRNA、体細胞の esiRNA という新しい RNA 群を発見した。また piRNA の生合成経路のモデルを提唱し、piRNA 合成と共役した Traffic jam 遺伝子の新規制御機構を明らかにした。ヒトの miRNA と Ago タンパク質の相互作用情報を収集し、新しい miRNA 多様化機構を見出した。</p> <p>○ヒトとマウスから 3000 を超えるセンス-アンチセンス RNA ペアの新規マイクロアレイを開発した。これを用いて正常組織と腫瘍部でセンス-アンチセンスペアの比率が著しく変動するペアを 60 個以上発見した。アンチセンス RNA 領域から産生される 60nt 程度の低分子 RNA 群を同定した。</p> <table border="1" data-bbox="470 1332 1390 1400"> <tr> <td>投稿論文</td> <td>「論文」 160件、「総説・解説記事等」 101件</td> </tr> <tr> <td>特許</td> <td>「出願済」 56件 (うち国際出願 26件)</td> </tr> </table>	投稿論文	「論文」 160件、「総説・解説記事等」 101件	特許	「出願済」 56件 (うち国際出願 26件)
投稿論文	「論文」 160件、「総説・解説記事等」 101件				
特許	「出願済」 56件 (うち国際出願 26件)				
<p>IV. 実用化の見通しについて</p>	<p>細胞分化誘導因子、遺伝子発現抑制因子、タンパク質翻訳調節因子等の機能を持った機能性 RNA を同定し、成果をすばやく特許化することにより、再生医療や RNA 医療、疾患診断、疾患治療等への応用を目指す。同時に、機能性 RNA を解析する手法及びツールの開発により、本分野における我が国の優位性の確立を目指す。</p> <p>さらに、本プロジェクトの成果は、医療分野以外のバイオ産業(農林水産産業、発酵工業、環境分野、食品分野、など)においても、大きな波及効果をもたらすと期待される。</p> <p><b>研究開発項目①「機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発」</b></p> <p>機能性 RNA データベース、次世代シーケンサー配列自動解析システムは、今後のエピゲノム解析などゲノム情報の医療・創薬への応用の情報基盤としての活用が予定されている。</p> <p><b>研究開発項目②「機能性 RNA 解析のための支援技術・ツールの開発」</b> (東京大学のグループ)</p> <p>・RNA マススペクトロメリー(RNA-MS)の開発に成功し、miRNA を診断分子として用いた血液診断や、核酸医薬の薬物動態技術に実用化のめどがたっている。さらに RNA-MS を支える情報解析技術に関しては著作物登録や国際出願を行っており、知財戦略の観点からも優位に立っている。</p>				

	<p>・RNAの全自動精製を実現した往復循環クロマトグラフィーは精製装置の開発に成功し、実用化に向けて検討を開始した。</p> <p>・これらの技術は国内の企業を中心に技術移転を行う予定である。とくに開発された世界最高レベルのRNA配列情報解析技術は、機能性RNA標的予測など核酸医薬の開発に必須の技術として活用を始めており、今後幅広く利用されることが期待される。</p> <p>(DNAチップ研究所)</p> <p>・高い検出感度を誇るmiRNAのマイクロアレイ技術の開発に成功しており、製品化のためのプロトタイプを生産している。</p> <p>(日本新薬のグループ)</p> <p>・高品質のRNAを安価に化学合成する方法を確立し、今年度中に「研究用試薬」としての製造販売を開始する計画である。将来的には、RNA医薬のGMP製造における「グローバルスタンダード」を目指している。</p> <p><b>研究開発項目③「機能性RNAの機能解析」</b></p> <p>(協和キリングループ)</p> <p>・マスト細胞の脱顆粒効率を制御するmiRNAや癌細胞増殖を制御するmiRNAを用いたアレルギー疾患や癌の治療技術の実用化に向けた研究が既に開始されている。</p> <p>(京大グループ)</p> <p>・iPS細胞作出効率を上昇させるmiRNAや間葉系幹細胞の分化を制御するmiRNAは、再生医療における高品質な有用細胞の作出技術への実用化が大いに期待される。</p> <p>(産総研グループ)</p> <p>・長鎖ncRNAやアンチセンスRNAでは、癌などの疾患の新規診断マーカーとして有望なものが得られている。</p>	
V. 評価に関する事項	事前評価	平成16年度実施。 担当部 バイオテクノロジー・医療技術開発部
	中間評価	H19年度 中間評価実施
	事後評価	H22年度 事後評価実施
VI. 基本計画に関する事項	作成時期	平成18年3月作成
	変更履歴	平成20年6月、イノベーションプログラム基本計画の制定により、「(1)研究開発の目的」の記載を改訂。また、人事異動に伴いプロジェクトリーダーの所属・役職を改訂。

用語	用語の解説
二次構造	RNA 配列上離れた位置にある2塩基が相補的な塩基対を形成することによって生じる RNA の構造。
動的計画法	直接的な計算では膨大な計算量を必要とする問題を部分問題に分割し、再帰式による計算によって効率的に解く方法の一般的名称。DNA 配列やアミノ酸配列の最適アラインメントを求める手法としても使われる。文字列の類似度だけを考慮したアラインメントは、配列長の2乗に比例したメモリと計算時間で求めることができるが、RNA 配列の二次構造を考慮した厳密なアラインメントでは、配列長の 6 乗の計算時間が必要とされ、解析の障害となってきた。
カーネル法	カーネル関数を用いた数理的手法の一般的名称。データを高次元に射影し、非線形データ構造を線形構造に置き換えることができる。サポート・ベクタ・マシンはカーネル法を用いた機械学習の手法である。
確率文脈自由文法	RNA の二次構造を解析するのに有用な確率的なモデルの一種。
構文解析	文章や文字列の構成要素間の文法的な関係を説明すること( <i>parse</i> )。構文解析は字句解析 ( <i>Lexical Analysis</i> ) とともに、プログラミング言語などの形式言語の解析に使用される。確率文法を用いた構文解析は DNA 配列からの遺伝子領域予測、RNA の二次構造予測に用いられている。
MEA 原理	アルゴリズムが予測するラベル系列が、正解のラベル系列にできるだけ一致するような予測を行う手法。
二次構造モチーフ	RNA の二次構造の特徴的な局所パターンのこと。
サポート・ベクタ・マシン	機械学習の手法の一種。近年最も強力な手法として急速に広まっている。
確率文脈自由文法	確率木文法と同様に、RNA の二次構造を解析するのに有用な確率的なモデルの一種。
シンテニー領域	異なる種の染色体間で、遺伝子が同じ順番で配置されている領域。
隠れマルコフモデル	線形の配列パターンをモデル化するための確率モデルの一種。主に、タンパクコーディング遺伝子発見に使われている。
多重配列アライメント (マルチプルアラインメント)	複数の配列(3 本以上)を相同性に基づいて整列させたもの。遺伝子の機能や進化の解析等に用いられる。マルチプルアラインメントを作成するための代表的なツールとして ClustalW がある。
Blastz プログラム	2本のゲノム配列の局所類似配列を検索するプログラム。
ペアワイズアラインメント	2本の配列を整列すること。
In silico	コンピューターを用いての意味。分子生物学などの実験は通常は、ウェット(wet)と呼ばれるように細胞や各種の生体分子を扱ったものであるが、それらの実験に関連するシミュレーションなどの計算などを指して in silico と呼ぶ。
マスペクトロメトリー法	質量分析法のこと。
RNA マスフィンガープリント法	RNA断片の質量数を、non-coding RNAデータベースに対して検索をかけることにより、RNAの配列を読むことなくRNA分子全体の配列を迅速に同定する方法。
サブフェムトモル(fmol)	モルの 10 のマイナス 16 乗のこと。分子数として約6億分子に相当する。
キャピラリーLC/ナノスプレ-イオン化法	内径0.1-0.2mmほどの細いキャピラリー管に樹脂を充填し、クロマトグラフィーを行うことで RNA 試料の微量分離が可能となる。溶離液を直接、高電圧を印加したスプレーヤーの先端部から噴霧することで RNA 分子をイオン化する方法。エレクトロスプレーイオン化法の一つである。
ランダムプライミング	ランダムな配列を持つオリゴDNAをプライマーとして、配列未知なRNAを



	鋳型にcDNAを作成する手法。
リンカーライゲーション	配列未知な RNA の配列を決定するために、末端に既知のオリゴ DNA や RNA を連結させること。
CID (collision-induced dissociation) 解析	イオン化した分子(ここでは RNA)を質量分析装置で稀ガスと衝突させることによって分解し、内部配列に由来する情報を取得する方法。RNA の de novo シーケンスに用いる。
蛍光相関分光法 (FCS)	レンズにより絞られたレーザー光の焦点領域の蛍光強度のゆらぎから蛍光分子の大きさと数を直接測定する方法。
核酸プローブ	検出したい RNA と特異的に結合する配列を持った核酸分子のこと。
アンチセンス遺伝子	蛋白を合成する mRNA の塩基配列(センス配列)に対して相補的な塩基配列をもつ遺伝子こと。
光ライゲーション反応	紫外光により隣接する核酸配列を連結する方法
MPEX反応	基板上での DNA 伸長反応による高感度に検出できる反応
ハイブリダイゼーション	相補的な塩基配列をもつ DNA(RNA)同士、もしくは DNA-RNA 間で 2 量体を形成すること。マイクロアレイ解析においては基盤上に配置されたプローブに、サンプル由来の標識された cDNA もしくは cRNA を作用させること。
マイクロ RNA	遺伝子の発現を抑制する働きがある機能性 RNA の一種
プライマー	DNA を酵素的に合成する際に使われる 20~30 塩基対の短い DNA 断片。
マイクロアレイ	様々な配列をもつ微量の DNA をスライドガラスやシリコン、ナイロン膜などの基板上に整列してのせ、固定化したものの総称。 サンプルの細胞と対象の組織細胞から別々に mRNA を抽出して、この RNA を別々の蛍光色素で標識し、cDNA (mRNA と相補的な塩基配列を持つ DNA) を合成する。これらの標識化された cDNA を混合してスライドガラス上にスポットして、固定化された遺伝子と結合させる操作(ハイブリダイゼーション)を用いて、遺伝子の発現量の比を検出する。DNA や RNA の遺伝子配列や発現量を解析することができ、生物が持つ遺伝子の僅かな違いを調べることができる。
DNA コンピューティング	DNA の分子生物学的性質を利用して計算する方法。
2' 位保護基	RNA は糖の隣接する 2' 位と 3' 位に水酸基を持ち、伸長反応中に 2' 位が反応しないように保護しておく必要がある。一般的には、TBDMs (t-butyl dimethylsilyl) 基が用いられているが、その嵩だかさが立体障害となり固相合成の収率を低下させる。一方、DNA は 2 位に水酸基がないため 2 位保護基が不要であり、RNA に比べて合成が容易であり、かつ 2 位の立体障害が軽減されているため固相合成の収率が高い。
モノマー	核酸の固相合成法(アミダイト法)における、P-N 結合(リン原子と窒素原子間の結合)を持つ A、T(U)、C、G 4 種類のホスホロアミダイト誘導体の重合体(ポリマー)の合成ユニットのこと。
レポーターアッセイ	ルシフェラーゼ(発光酵素)遺伝子をレポーターとし、1 種類の試薬で同時発光させることで転写活性を測定すること。
間葉系幹細胞	間葉系幹細胞とは、骨髄液や脂肪組織などの中に少量存在する未分化の細胞で、筋肉、骨、脂肪など種類の異なる複数の細胞に分化できる能力を持ち、かつ自己複製の能力も持つ細胞のことを言う。
マスト細胞	喘息、アトピー性皮膚炎などの即時型アレルギー反応を誘起する細胞のこと。肥満細胞とも言う。膜表面に IgE の受容体を有し、ヒスタミンなどを含有する顆粒を含むことが特徴。IgE の刺激などで起こる脱顆粒によ

	り、顆粒内のヒスタミンや化学伝達物質が放出されることで、肺の気管支収縮や粘膜浮腫などのアレルギー症状が起こる。
癌細胞株	ヒトやマウス、ラットの腫瘍組織から単離され、株化された細胞のこと。不死化しており無限増殖能を有する、足場非依存性の増殖を示す等の特徴を示す。
MPSS 法	Massively Parallel Signature Sequencing の略で、メッセンジャーRNAに由来する配列の一部を数十万個、一度に取り出して各遺伝子の発現量を測定する方法。
アポトーシス	多細胞生物の体を構成する細胞の死に方の一種で、個体をより良い状態に保つために積極的に引き起こされる、管理・調節された細胞の自殺(プログラムされた細胞死)のこと。
カリオタイプ	染色体の核型
ヒト臨床腫瘍サンプル	癌患者由来の組織及びそこから単離された DNA、RNA、タンパク質等のこと。固形癌の場合は手術時に切除した腫瘍組織とそれに付随した隣接正常組織が利用される。血液系癌の場合には採血により取得される。
ノックアウトマウス	機能抑制系の有力な方法として、マウス個体において機能を調べたい遺伝子を完全に欠失させる手段がある。そのようにして作成された遺伝子改変マウスのこと。遺伝子の機能解析において極めて強力な手法であるが、手間・技術・時間が必要とされる。
ノックダウン	RNA を導入(発現)することにより、遺伝子発現を抑制すること。
脆弱 X 症候群(fragile X syndrome: FRA X)	精神遅滞・巨大睪丸・染色体検査による脆弱 X 所見を主徴とする奇形症候群。原因は、X 染色体上の FMR1 遺伝子の機能不全とされる。
RNAi	RNA 干渉ともいう。細胞に二本鎖 RNA を導入した場合、それと同じ配列をもつ遺伝子の発現(タンパク質の合成)を抑制する現象のこと。真核生物において、標的遺伝子(mRNA)を破壊(分解)することにより遺伝子発現を制御し、ウイルスに対する防御反応と類似した現象であり、遺伝子解析の有効な方法として、基礎研究から将来の医療分野への応用として期待されている。
モノクローナル抗体	認識部位が単一の特異性の高い抗体のこと。
RISC	RNA 誘導サイレンシング複合体(RNA-induced silencing complex)。RNA 干渉における遺伝子の不活性化やマイクロ RNA の機能発現において重要な役割を果たす。
センス-アンチセンス	通常アンチセンス遺伝子は、タンパク質をコードする遺伝子が転写された鎖(センス鎖)と相補的な配列を有する(反対鎖から転写される RNA をコードするような)遺伝子を指すが、センス-アンチセンスの関係は相対的であるため、これらを一括してセンス-アンチセンスと呼ぶ。

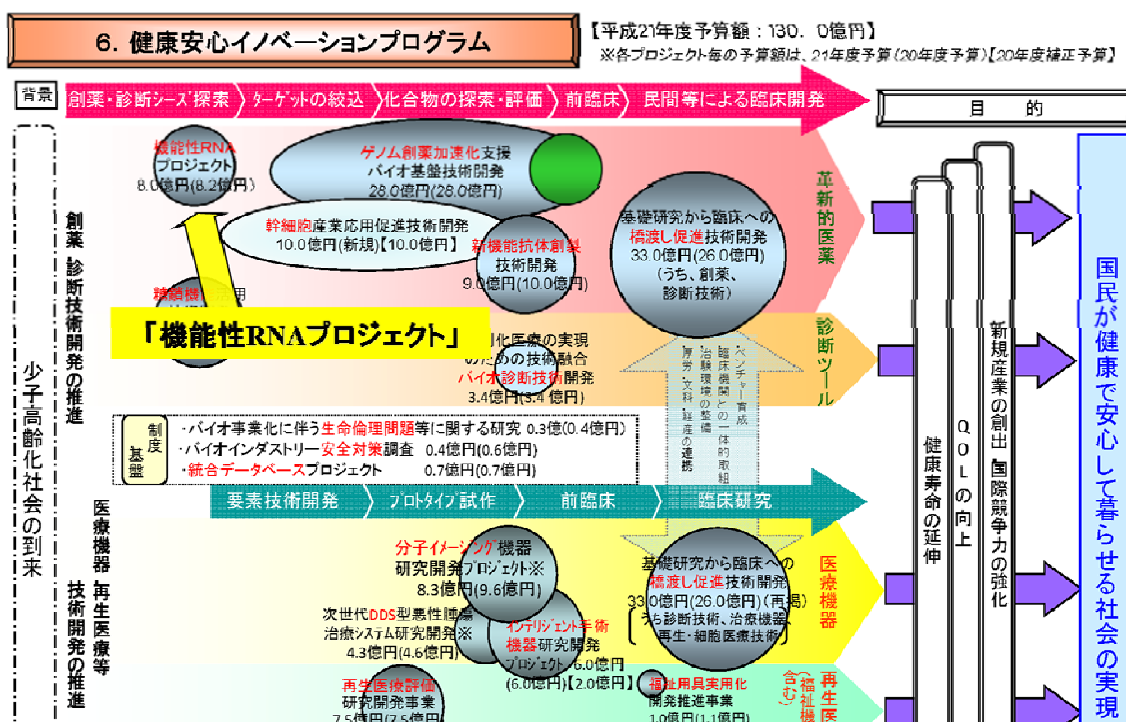
## I. 事業の位置付け・必要性について

### 1. NEDOの関与の必要性・制度への適合性

#### 1.1 NEDOが関与することの意義

Non-coding RNA(ncRNA:非コードRNA)は、タンパク質をコードする mRNA 以外の RNA の総称である。細胞内に存在する全転写産物には相当量の ncRNA が存在することが推測されており、細胞機能を理解する上でそれらの機能解析は欠くことのできない重要課題である。昨今、このように機能を有すると推定される ncRNA、即ち、機能性 RNA による遺伝子発現調節機構が生命の発生や細胞の分化に重要な役割を果たすことが明らかになりつつある。

NEDO技術開発機構では、すでに研究開発が着手されている遺伝子、タンパク質、糖鎖等の生体分子の機能・構造・ネットワーク解析に加え、新たなポストゲノム研究の展開として機能性 RNA の解析を行うとともに、それらの研究を強力に推進するためのバイオツールやバイオインフォマティクスの開発、成果を高度に利用するためのデータベース整備や先端技術を応用した高度医療機器開発等により、テーラーメイド医療・予防医療・再生医療等の実現や画期的な新薬の開発、医療機器、福祉機器等の開発・実用化を促進することによって健康寿命を延伸し、今後、世界に類を見ない少子高齢化社会を迎える我が国において、国民が健康で安心して暮らせる社会の実現を目指すことを目的とする「健康安心イノベーションプログラム」の一環として本プロジェクトを遂行することとした(「添付書類1. 健康安心プログラムイノベーションプログラム基本計画」を参照)。



生体における機能性RNAの機能を解析し、その機能の抑制や促進による生体応答

を総合的に把握する基礎的・基盤的研究を強力に進めることは、再生医療や疾患治療等の実現に資する重要な課題である。本プロジェクトは基礎的・基盤的研究であるが、開発する技術や研究開発過程で取得される各種のデータは、画期的な新薬の開発をはじめとした幅広い分野に直結している。この分野における国際的社会的背景と、開発すべき技術の多様性・先端性・緊急性を鑑みれば、個別企業の経営資源と開発力とではとても賄える規模の研究開発では無く、また開発技術・成果は国内バイオ・IT産業の基盤を担うものである。このような研究開発を達成するには、NEDOの関与が必要であり、我が国の企業の研究開発事業等をサポートすることへ大きな意義がある。

## 1. 2 費用対効果

本プロジェクトにおいては、バイオインフォマティクスの活用による機能性 RNA を推定する技術の開発、機能性 RNA 解析のための新たな支援技術・ツールの開発、及び機能性 RNA の機能解析を実施する。これらを通して、その遺伝子発現調節機構を明らかにすることにより、細胞分化誘導因子、遺伝子発現抑制因子、タンパク質翻訳調節因子等の機能を持った機能性 RNA を同定し、バイオ分野における新たな基盤形成を目指す。これにより、再生医療や RNA 医療、遺伝子治療、疾患治療等、産業応用の促進、新産業の創出が期待される。また同時に、機能性 RNA を解析する手法及びツールの開発により、機能性 RNA 解析研究を飛躍的に進めることで、本分野における我が国の優位性の確立を目指す。さらに本プロジェクトにおける成果を素早く特許化することにより、将来的な医療・診断分野における新産業の創出に貢献するなど、その実施効果は大いに期待できる。

## 2. 事業の背景・目的・位置づけ

近年の研究により、我々哺乳類を含む高等生物の細胞中には、従来のタンパク質をコードする RNA とは異なり、タンパク質をコードしていないにもかかわらず転写される non-coding RNA(ncRNA:非コード RNA)が多数存在することが明らかになった。これらは、機能性 RNA として、発生や細胞分化の過程において重要な役割を果たしており、癌や糖尿病などの疾患の発生にも深く関わっているものと考えられている。

本プロジェクトでは、バイオインフォマティクスの活用による機能性 RNA を推定する技術の開発、機能性 RNA 解析のための支援技術・ツールの開発、及び機能性 RNA の機能解析を行うことにより、本研究分野における我が国の優位性の確立を目指す。また、本プロジェクトにおける成果を素早く特許化することにより、将来的な医療・診断分野における新産業の創出に貢献する。

## II. 研究開発マネジメントについて

### 1. 事業の研究開発目標

#### 1.1 研究開発目標

最終目標(平成21年度末)

機能性 RNA 候補をゲノム上から推測するバイオインフォマティクス技術を確立する。また、機能性 RNA を高感度、定量的かつ網羅的に捉える新しい方法、手法の確立と、機能性 RNA をゲノムワイドに解析するためのツールを確立する。これらのツールを活用して、機能性 RNA の機能解明に取り組む。また、ヒト疾患に関連する機能性 RNA 及び発生・分化などをはじめ細胞機能に重要な働きを示す数十個の機能性 RNA 候補の機能解析を行い、医薬品開発や再生医療等に有用な基盤知見の取得や、基盤技術の構築を目指す。(「添付書類 2. プロジェクト基本計画」を参照)

#### 1.2 目標設定の理由

三つの研究開発項目である、①機能性RNAの探索・解析のためのバイオインフォマティクス技術の開発、②機能性RNA解析のための支援技術・ツールの開発、③機能性RNAの機能の解明、に沿った目標設定であり、実用化進展を踏まえて幅広く捉えた目標内容でもある。世界的にも競争の激しい分野であり、確実な基盤知見の取得、基盤技術の構築を目指し、産業化への出口イメージをも踏まえた目標となっている。

## 2. 事業の研究開発計画

### 2.1 研究開発項目①「機能性RNAの探索・解析のためのバイオインフォマティクス技術の開発」

【達成目標】(平成21年度末)

バイオインフォマティクス技術を活用して新規の機能性RNA候補を網羅的に予測し、機能性RNAデータベースを構築する。(「添付書類 2. プロジェクト基本計画」を参照)

【研究の必要性】

膨大な塩基配列データから、機能を持つ RNA 配列部分をすべて実験によって発見することは困難であるため、ゲノム配列・cDNA 配列などから機能性 RNA を網羅的に発見する情報技術が必要である。また、配列類似性と同時に二次構造が重要な機能性 RNA の情報解析には、既存の配列相同性解析だけでは不十分であり、機能性 RNA に特化したバイオインフォマティクスの研究開発が必要である。さらに、本プロジェクト等で得られた情報を広く一般に効果的に活用するために、そのデータベースを構築することが必要である。

【研究開発の具体的内容】

機能性RNA配列を予測するバイオインフォマティクス手法の開発を実施する。また、このような技術等を活用し、ゲノム配列からの機能性RNAの網羅的な予測を行い、さらに

機能性RNAのデータベースを構築する。

①-1. 機能性RNAに特化したバイオインフォマティクス技術の開発

二次構造の類似性を考慮できるRNA配列に特化した、配列比較・検索手法を開発する。また、RNA配列の整列、配列群からの二次構造予測、機能推定を行う情報技術を開発する。また、機能性RNAをターゲットとするマイクロアレイ情報の解析技術を開発する。

①-2. ゲノム配列からの機能性RNAの網羅的予測

機能性RNAに特化したバイオインフォマティクス技術を活用し、ゲノム配列全体から機能性RNAを網羅的に予測する。

①-3. 機能性RNAデータベースの構築

機能性 RNA データを広く一般に効果的に活用するため、機能性RNAデータベースを構築する。

## 2.2 研究開発項目②「機能性 RNA 解析のための支援技術・ツールの開発」

【達成目標】(平成21年度末)

機能性 RNA をハイスループット、高感度、網羅的に解析できるバイオツールを開発する。また、超微量な機能性RNAを高感度(サブフェムトモルオーダー)で直接測定することが可能な方法・手法等の開発。

【研究の必要性】

機能性RNAを単離、精製してin vitro実験で解析する手法・ツールは、開発されつつあるが、感度面と安定性において課題が残っている。特に機能性RNAは発現が極微量でありながら、発生や分化などの生命現象で重要な機能を持つことが知られており、機能性RNAの発現量や発現変動を捉えるためには高感度なバイオツール開発が不可欠である。さらに、機能性RNA解析を加速するため、in vitro実験においてハイスループットに解析する手法・ツール開発が必要である。

【研究開発の具体的内容】

機能性RNAをハイスループットに、網羅的に、かつ定量的に再現性をもって解析できるバイオツールの開発を行う。また、発生・分化・疾患等により変動する機能性RNAの発現変動を網羅的に解析する技術の開発を行う。さらに、RNAはタンパク質と比較して更に微量であるため、測定を極限まで高感度化をはかることで、細胞から抽出した超微量な機能性RNAを高感度で直接測定する技術・手法を開発する。

②-1. RNA のマススペクトロメトリー法の開発

②-2. 機能性 RNA の高感度検出システムの開発

②-3. RNA の新規合成基盤技術開発と化学分子設計

## 2.3 研究開発項目③「機能性 RNA の機能解析」

【達成目標】(平成21年度末)

ヒト疾患に関連する機能性 RNA 及び発生・分化などをはじめ細胞機能に重要な働きを示す数十個の機能性 RNA 候補の機能解析を行い、医薬品開発や再生医療等に有用な基盤知見の取得や、基盤技術の構築を目指す。

#### 【研究の必要性】

種々の機能性RNAが、どの遺伝子を標的として、どういった機構で作用することによって発現制御を行っているか、また、機能性RNA自身の発現・機能調節はどの様に行われているか、等に関する知見は極めて少ないのが現状であり、これらの機能性RNAに関する発現・機能制御機構の解明が非常に重要となっている。さらに、疾患に関連した機能性RNAを同定、解析し、診断マーカーとしての展開、及び疾患治療や再生医療等を目指した医薬品の開発基盤を確立することも必要である。

#### 【研究開発の具体的内容】

ヒト疾患に関連する機能性RNAの同定とその機能解析、機能性RNAの作用機序の解明、また発生・分化など重要な働きを示す機能性RNAの同定とその機能解析を行う。

##### ③-1. ヒト疾患に関連する機能性RNAの迅速で高効率な同定

ヒト疾患に関連する機能性RNA候補を選抜し、疾患細胞株、疾患モデル動物等での評価を実施し、疾患で変異あるいは発現変動が見られる機能性RNAを同定する。これらの各種疾患関連機能性RNAに対して、*in vitro*、及び*in vivo*系で各種解析手法を駆使して機能解析を行う。特定された機能性RNAについては、核酸関連低分子化合物等を用いて機能性RNAの発現・機能を制御するための研究開発を行う。

(1) ヒト疾患に関連する機能性 RNA の同定と機能解析

(2) 機能性 RNA 解析に基づくゲノム医学研究

##### ③-2. 機能性RNAに関する基盤的知見の獲得とそれを基にした機能性RNA同定

発生・分化等におけるダイナミックな機能性RNAの発現変動解析、同定を行い、新規な機能を有する機能性RNAの発見を目指す。さらに機能性RNAの生体細胞内での生合成経路、局在、結合パートナーの同定を通して、機能性RNAの機能を解明し、その機能を人為的に制御するシステムを構築する。

(1) 機能性 ncRNA の多面的選別法の確立と機能解明

(2) マイクロ RNA の作用機構の解明

(3) アンチセンス RNA の機能解析

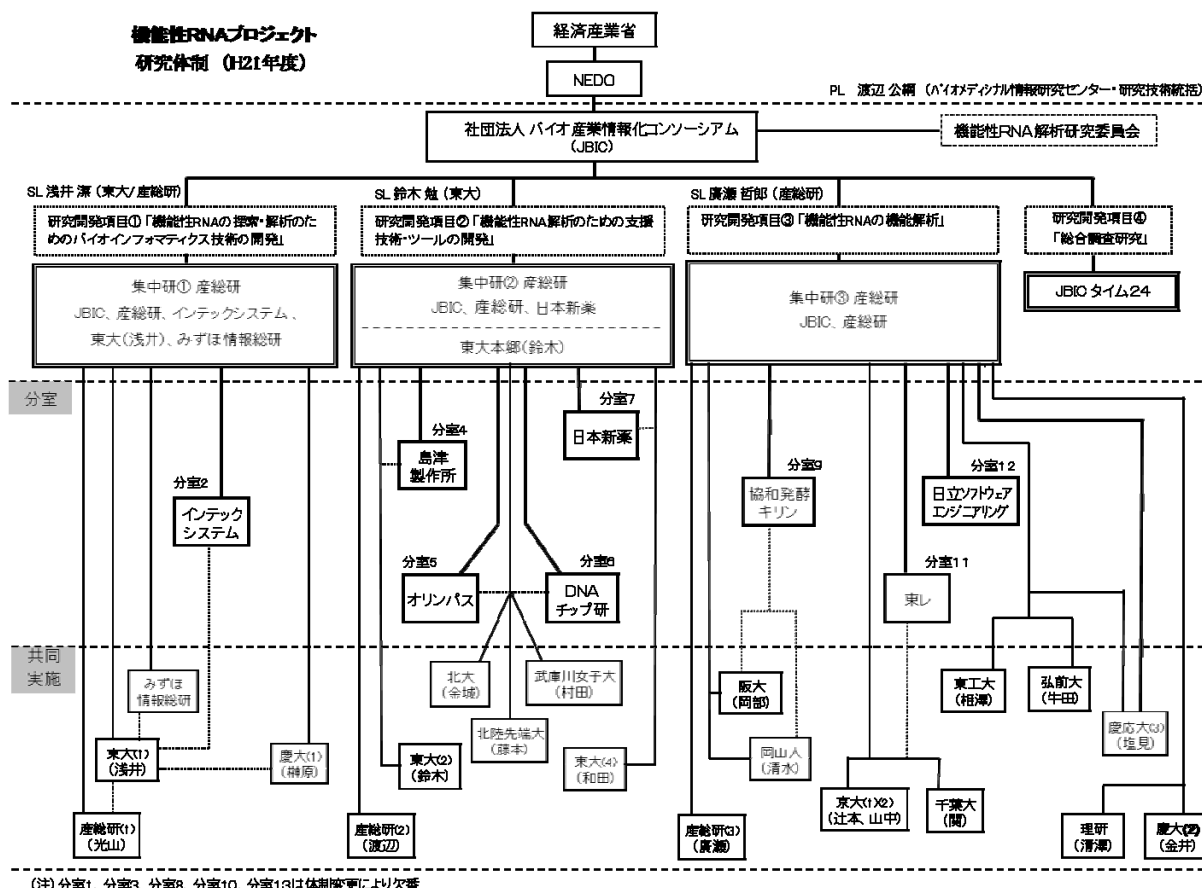
### 3. 研究開発実施の事業体制

#### 3.1 事業体制

本研究開発は、平成17年度に経済産業省産業技術環境局研究開発課及び製造産業局生物化学産業課において基本計画を策定し本事業を実施した。平成18年度～平成21年度は、独立行政法人新エネルギー・産業技術総合開発機構(以下「NEDO技術開発機構」という)において本事業を実施した。この際に「委託契約事務の取扱いに関する機構達

(平成15年機構達第8号)第14条「公募によらない場合の適用基準」に基づき、NEDO技術開発機構において公募による研究開発実施者の選定は行わなかった。

また、独立行政法人産業技術総合研究所生物情報解析研究センター 渡辺公綱 研究技術統括を研究開発責任者(以下「プロジェクトリーダー」という)とし、その下に研究者を可能な限り結集して効果的な研究開発を実施した。[研究体制図を参照]



### 3.2 研究開発の運営管理

研究開発全体の管理・執行に責任を有するNEDO技術開発機構は、経済産業省、プロジェクトリーダー(PL)である渡辺公綱、及び委託先である社団法人バイオ産業情報化コンソーシアム(JBIC)と密接な関係を維持しつつ、「健康安心イノベーションプログラム」の目的及び目標、並びに本事業の目的及び目標に照らして運営管理を実施した。

具体的には、渡辺公綱PLを委員長とした「機能性RNA解析研究推進委員会」を組織し、原則月1回「サブリーダー会議」を開催した。各分室・共同実施先からは毎月「月報」を提出してもらい、サブリーダー会議においてレビューを行った。本委員会は「お台場機能性RNAコロキウム」(研究進捗討議会)を年間に8回程度開催し、各研究開発項目間における情報交換の場とした。「研究進捗報告会」(全体会議)を原則年1回開催し、外部委員を招集し、プロジェクト方針についての助言を得た。さらに、平成19年1月には「研究進捗状況



ヒアリング」を2日間に渡って開催し、その結果を次年度の研究実施計画に反映させた。それぞれの開催実績は以下の通りである。

#### ○月報の提出:

各グループ(分室、共同実施先)の業務管理者または主要研究員が、翌月の5日までに、A4版1ページに進捗状況を記載し、SPLにメールで提出した。(秘密情報を含む場合はパスワード保護をかけるか、または郵送とした) 下記サブリーダー会議において、各SPLがお互い担当テーマの状況を紹介した。多岐にわたるプロジェクト各テーマの進捗概要を把握するのに非常に有効であった。

#### ○サブリーダー会議の開催:

プロジェクトリーダー(PL)、3名のサブリーダー(SPL)およびJBICのプロジェクト担当者が原則毎月会合し、プロジェクトの運営について協議・決定した。

開催日時:

H17(2005)年度(6回)

9/9, 11/18, 12/12, 1/18, 2/1, 3/23

H18(2006)年度(11回)

4/26, 5/24, 6/28, 7/24, 9/19, 10/25, 11/27, 12/19, 1/24, 1/29, 3/28

H19(2007)年度(4回)

4/11, 4/18, 5/17, 6/12, 8/24, 9/18, 10/26, 11/28, 12/25

H20(2008)年度(8回)

5/7, 6/18, 7/18, 9/10, 11/14, 12/18, 1/22, 2/20,

H21(2009)年度(6回)

5/21, 6.26, 8/19, 10/13, 11/18, 12/25

会議メンバー:

氏名		所属・役職
渡辺 公綱	PL	産総研 生物情報解析研究センター・研究技術統括
浅井 潔	SPL①	東京大学 大学院新領域創成科学研究科・教授 産総研 生命情報工学研究センター・センター長
鈴木 勉	SPL②	東京大学大学院工学系研究科・教授
廣瀬 哲郎	SPL③	産総研 生物情報解析研究センター・チーム長
中川 智	委託先	JBIC 研究開発本部・担当部長

#### ○お台場機能性RNAコロキウム(研究進捗討議会)の開催:

原則月1回の頻度で、集中研(お台場)において開催した。3名の研究員が研究進捗状況を報告し、ディスカッションを行った。時により、学会参加報告会とした。毎回40名程度の参加があり、活発な情報交換の場となった。

### ○研究進捗報告会(全体会議)の開催:

2005年10月3日 13:00~18:00 キックオフ・ミーティング

各グループの研究計画紹介。参加者101名。

2006年2月22日 13:00~18:00 平成17年度研究進捗報告会

各グループの研究報告。外部評価委員2名。参加者87名。

2007年2月23日 10:00~18:00 平成18年度研究進捗報告会

各グループの研究報告。外部評価委員4名。参加者97名。

2008年2月13日 10:00~18:00 平成19年度研究進捗報告会

各グループの研究報告。外部評価委員3名。参加者89名。

2009年2月13日 10:00~18:00 平成20年度研究進捗報告会

各グループの研究報告。外部評価委員4名。参加者81名。

2010年1月22日 10:00~18:00 平成21年度研究進捗報告会

各グループの研究報告。外部評価委員3名。参加者81名。

### ●研究進捗状況ヒアリングの実施:(自主点検)

H19(2007)年1月23日および24日(両日とも10:00~18:00)において「機能性RNA解析研究推進委員会」が各グループ(集中研、分室、共同実施先)の研究進捗状況を個別にヒアリングし、評価を行なった。この評価結果に基づき、平成19年度の研究実施計画(体制、予算配分)を決定した。

#### お台場機能性RNAコロキウム開催一覧

第1回 2006年11月29日(火)16:30~19:00 参加者40名

金 大真(産総研)「RNAのバイオインフォマティクス研究の課題と展望」

堀 邦夫(ハスジーン)「蛍光相関分光法による核酸分子の挙動測定」

廣瀬 哲郎(産総研)「mRNA型 ncRNA の研究法についての考察」

第2回 2006年12月12日(月)15:00~17:00 参加者42名

北川 英俊(日本新薬)「新しい保護基 CEM を用いた RNA の化学合成法の開発」

相澤 康則(東工大)「ヒト間葉系幹細胞を用いた機能性 RNA 研究の展望」

牛田 千里(弘前大)「線虫を材料として用いた機能性 RNA 研究の利点」

第3回 2006年1月18日(水)13:00~15:00 参加者44名

浅井 潔(東大)「革新的なRNA配列情報解析技術の開発」

村田 成範(DNAチップ研)「MPEX法を用いた RNA の微量検出法の構築」

森川 實(ジェノタイプファーマ)「尋常性乾癬に關与する機能性 RNA」

第4回 2006年3月23日(木)13:30~15:30 参加者41名

齋藤 輪太郎(慶応大)「センス・アンチセンス RNA のバイオインフォマティクス」

清澤 秀孔(理研)「センス・アンチセンス RNA の網羅的発現解析」

塩見 美喜子(徳島大)「マイクロ RNA の機能発現に關与するタンパク質因子の解析」

第5回 2006年4月20日(木)13:30~16:00 参加者48名

榊原 康文(慶応大)「確率文法とカーネル法による RNA 配列情報解析技術の開発」

鈴木 勉(東大)「RNAのマススペクトロメトリー法の開発」

関 直彦(千葉大)「ヒト腫瘍組織を用いた機能性 RNA の発現解析」

第6回 2006年5月29日(木)13:30~15:30 参加者47名

浜田 道昭(みずほ情報総研)「RNAの二次構造に着目した革新的な配列解析アルゴリズムの開発」

斎藤 博英(京大)「機能性モチーフとその結合蛋白質を用いたリボスイッチの構築」

- 吉田 哲郎 (協和発酵) 「ユニークな培養細胞系からの新規マイクロ RNA のクローニングと機能解析」  
 第7回 2006年6月28日(水)13:00~15:00 参加者41名  
 金 大真 (産総研 CBRC) 「データベース関連の最近の状況について」  
 長尾 一生 (JBIC 研究員) 「機能性 RNA を in vivo で計測するシステム及び高感度検出システムの開発」  
 井手上 賢 (JBIC 研究員) 「H-inv 登録 Non-coding RNA 候補の細胞内局在と RNA 制御因子の作用について」  
 第8回 2006年7月19日(木)13:00~15:30 参加者41名  
 \* 学会参加報告  
 加藤 敬行 RNA2006 Annual Meeting of RNA Society (Seattle, 米国)  
 渋谷 利治 RNA2006 Annual Meeting of RNA Society (Seattle, 米国)  
 廣瀬 哲郎 RNA2006 Annual Meeting of RNA Society (Seattle, 米国) + 日本 RNA 学会年会  
 井手上 賢 IUBMB International Congress (京都) + 日本 RNA 学会年会  
 長尾 一生 IUBMB International Congress (京都)  
 相澤 康則 Cold Spring Harbor Laboratories Symposium "Regulatory RNA" (New York, 米国) + 日本 RNA 学会年会  
 牛田 千里 Cold Spring Harbor Laboratories Symposium "Regulatory RNA" (New York, 米国) + 日本 RNA 学会年会  
 \* 番外  
 金 大真 (産総研) 「バイオインフォマティクス・ツールの解説」  
 第9回 2006年9月19日(火)15:30~17:30 参加者38名  
 佐藤 健吾 (JBIC 研究員) 「Benasque 会議(機能性 RNA ワークショップ、スペイン)の報告」  
 植竹 弘一 (日本新薬) 「RNA の新規合成基盤技術開発と化学分子設計」  
 相澤 康則 (東工大) 「ヒト間葉系幹細胞(hMSC)の分化に関わる ncRNA の探索」  
 第10回 2006年10月23日(月)15:30~17:30 参加者26名  
 村田 成範 (DNA チップ研究所) 「マイクロレイ上での酵素反応系を用いた短鎖 RNA の微量検出法の構築」  
 田中 正史 (ジェノタイプファーム) 「尋常性乾癬感受性領域に発現する ncRNAs」  
 牛田 千里(弘前大) 「Movement of *C. elegans* small ncRNAs in oogenesis, fertilization and early embryonic cell division」  
 第11回 2006年11月27日(月)15:00~17:00 参加者45名  
 塩見 美喜子 (徳島大) 「ショウジョウバエ *Aubergine* と rasiRNA を介して起こる RNA silencing」  
 吉田 哲郎 (協和発酵) 「ヒトマスト細胞からの miRNA の単離と発現解析」  
 加藤 敬行 (東大) 「RNAi 活性に影響を及ぼす因子の同定と高効率 siRNA 設計法の開発」  
 第12回 2006年12月19日(火)15:30~17:30 参加者36名  
 井上 丹 (京大) 「ncRNA の機能制御システムの開発」  
 寺澤 和哉 (京大) 「分化に関わる miRNA の同定とその機能解析」  
 清澤 秀孔 (理研) 「ヒト組織、及び腫瘍組織におけるセンス-アンチセンス RNA の発現解析」  
 第13回 2007年4月18日(水)15:00~17:00 参加者43名  
 \* 最近の国際学会レポート  
 佐藤 健吾 (CBRC): BIRD 2007  
 加藤 敬行 (東大): キーストーンシンポジウム  
 櫻井 雅之 (東大): ゴードン会議  
 長尾 一生 (JBIRC): 生物物理学会  
 青木 一真 (JBIRC): RNA2006Izu  
 井手 上賢 (JBIRC): RNA2006Izu  
 第14回 2007年5月17日(木)15:30~17:30 参加者39名  
 木立 尚孝 (CBRC) 「RNA 配列多重アラインメントツール Murlet」  
 上田 宏生 (東大) 「RNA マスフィンガープリント法の開発」  
 佐々木 保典 (BIRC) 「ヒト機能性 RNA の基盤的特性の発見」  
 第15回 2007年7月23日(月)15:30~17:30  
 菊池 (東工大) 「分化誘導系によるノンコーディング遺伝子の抽出と発現解析」  
 牛田 (弘前大) 「組織特異的発現を示す新規 ncRNA について」  
 陶山 (東大) 「アトモル以下の多種類の RNA の同時定量を逆転写をせずに行うフォト DEAN 法」  
 第16回 2007年9月18日(火)15:30~17:30  
 浅井潔 (東大、産総研) 「RNA 配列情報解析ツールの現状と今後の課題」  
 北川英俊 (日本新薬) 「CEM 法を用いる長鎖 RNA 合成と修飾核酸合成への取り組み」  
 沼田興治 (理研) 「人工逆鎖プローブをもちいた新規内在性アンチセンス RNA の予測に向けた試み」  
 第17回 2007年10月26日(金)13:00~15:00  
 村田 成範 (DNA チップ研究所) 「miRNA 検出用マイクロレイツールの開発経過報告と今後の展開」  
 塩見 美喜子 (徳島大学) 「内在性ヒト Argonaute に結合する small RNA の解析と機能相違」

- 中野 春男 (協和発酵)「癌細胞の増殖・細胞死に関するマイクロRNAの同定と機能解析」  
 第18回2007年11月28日(水)13:00～15:00  
 金大真(産総研)「機能性RNAデータベースの構築とカスタムマイクロレイトを用いた新規RNA遺伝子候補領域の発現解析」  
 宮内健常(東大)「往復循環クロマトグラフィーによる機能性RNAの単離精製」  
 小柳三千代(京大)「マウス体細胞の初期化に関わる microRNA の探索とその機能解析」  
 第19回2007年12月25日(水)15:30～17:30  
 寺井 悟朗(インテックW&G/JBIC)「パラログ解析に基づく機能性RNA発見」  
 藤本 健造(北陸先端大)「RNA解析に向けた光ライゲーション技術の開発」  
 青木 一真(JBIRC)「ヒト機能性ncRNAの機能解析に向けた組織特異性を指標としたアプローチ」  
 第20回2008年4月18日(金)15:15～17:15  
 坂口 裕理子(東大)「RNAマスペクトロメリーの開発」  
 井手上 賢(BIRC)「核内機能性RNAの機能解析」  
 相澤 康則(東工大)「ヒト幹細胞関連ノンコーディングRNA遺伝子の機能分類」  
 第21回2008年11月17日(月)15:30～17:30  
 佐藤 健吾(産総研CBRC/JBIC)「カーネル法による機能性RNA遺伝子の探索」  
 櫻井 雅之(東大工/JBIC)「ヒト転写産物におけるイノシン化部位の網羅的探索と機能解析」  
 清澤 秀孔(理研)「アンチセンス転写とエピジェネティックな遺伝子発現制御」

#### 4 研究開発成果の実用化に向けたマネジメント

##### 4.1 知財戦略

###### 【方針】

1. 本プロジェクト成果において、実用化の核となる基盤技術をコア技術として、その知財を重点的に確保する。
2. 個別のテーマ(機能性RNAの探索・発見など)については参画企業の戦略(医薬開発戦略等)を尊重。

###### 【成果】

区分 年度	特許出願		主なコア技術特許
	国内	国外	
H17	1件	0件	①RNAの二次構造予測技術(アルゴリズム)
H18	8件	2件	
H19	7件	6件	②RNAの質量分析技術
H20	9件	11件	
H21	5件	7件	③RNAの生物学的ノックダウン分析技術
小計	30	26	
計	56		③iPS細胞作成技術

##### 4.2 成果実用化を推進するマネジメントの取り組み

###### 【方針】

本プロジェクト成果を他のNEDOプロジェクトに紹介し組合せ、実用化に向けたシナジー効果を模索する。

###### 【成果】

研究開発項目②における「RNAの質量分析技術」を核酸医薬による遺伝子治療に関する他のNEDOプロジェクトに紹介し、モデル動物の血中での核酸医薬候補の分析を実

施し成功。該遺伝子治療の有用性の一端が証明され、治療推進の根拠となった。

## 5. 情勢変化への対応

### 5.1 研究テーマおよび研究体制の見直しと強化・発展

○「研究進捗状況ヒアリング」(H19年1月)における見直し(自主点検の実施)

- ・テーマ③-1-(2)「尋常性乾癬に関わる RNA の機能解析」については、当初予定していた臨床検体の入手ができておらず、研究が大幅に遅れてしまい、今後の改善の目処も立たないため、平成18年度をもって本テーマは中止した。この結果、ジェノダイブファーマ株式会社(分室10)と東海大学(共同実施先)を研究実施体制から外した。
- ・テーマ②-2「機能性 RNA を in vivo で計測するシステムの開発」については、標識した RNA を細胞内に取り込ませる技術の確立に手間取り、研究が大幅に遅れてしまった。そこで、当面は、本テーマは中断することとし、計画通り進捗しているテーマ②-3「機能性 RNA の高感度検出システムの開発」の方に研究資源を集中することとした。その他のテーマについても、評価結果を平成19年度の予算配分に反映させた。

◎加速予算を投入(H19年4月)し研究を推進

- ・テーマ②-1「RNA のマスペクトロメトリー(質量分析技術)の開発」は、本プロジェクトを支える重要なツール開発であり、世界的に見ても日本発のオリジナルな技術である。実施計画をはるかに上回るペースで開発が進んでいることから、NEDO技術開発機構が独立行政法人に移行する際に新設された、著しい成果を挙げているプロジェクトに対して追加的な資金を投入する「加速財源制度」を活用し、平成19(2007)年4月にフーリエ変換型質量分析装置を導入した。本装置は、2005年にサーモエレクトロン社が発売したもので、質量分解能が飛躍的に高く、創薬分野や精密化学分野では必要不可欠な装置としての定評を得ている。RNA 研究に応用されるのも時間の問題と思われる、本プロジェクトが築き上げた優位性を確保するために、いち早く本装置を導入した。

○中間評価(H19年4月)における見直し

「5.2 中間評価の受審と対応」を参照)

○最終年度にむけた見直し【H21年4月】

テーマ①-2「ゲノム配列からの機能性 RNA の網羅的予測」における「比較ゲノムによるゲノム相同領域の抽出」(三菱総研担当)については一定成果が得られ、抽出した領域の解析に研究開発の重点が移ったため、プロジェクト途中で終了した。

### 5.2 中間評価の受審と対応

#### 5.2-1 中間評価

NEDO技術開発機構は、技術的及び政策的観点から、研究開発の意義、目標達成度、成果の技術的意義並びに将来の産業への波及効果について、外部有識者による研究開発の中間評価を平成19年度に以下のとおり実施し、その結果を踏まえ、プロジェクトの見

直しを実施した。

#### 【中間評価】

##### ①中間評価分科会

日時:平成19年7月4日(水) 13:00～17:30

場所:日本科学未来館 7階 会議室2

##### ②評価手法: 外部評価

##### ③中間評価委員:

分科会長	大石 道夫	かずさDNA研究所	理事長兼 所長
分科会長 代理	饗場 弘二	名古屋大学 大学院 理学研究科 生命理学専攻	教授
委員	中井 謙太	東京大学 医科学研究所 ヒトゲノム解析 センター 機能解析イン・シリコ分野	教授
委員	名取 幸和	東京工業大学 大学院 理工学研究科 コアリッションセンター機能体	特任教授
委員	西島 和三	持田製薬株式会社 医薬開発本部 開発企画推進部	主事
委員	横山 勇生	日経BP社 日経メディカル別冊	編集委員

#### 5.2-2 中間評価結果

##### 肯定的な評価ポイント

機能性RNAのバイオインフォマティクスによる予測技術、検出・解析技術、および生体機能と作用機構の解明の目標を掲げ、順調に研究を進めている。興味深い結果も創出しており、後半は更なる成果が期待される。機能性RNA研究は世界的な競争の状態にあり、将来のバイオ産業の新たな柱を生み出す可能性のある技術である。基盤研究の構築・確立は我が国の民間企業のみでは到底達成できない分野であり、本プロジェクトを実施する意義は高い。

対応を要する評価ポイント:

一方、欧米においてその限界が明確となったり、逆に最終目標のハードルが高い分野もあるので、世界の趨勢をよく見極め、テーマの絞り込み、予算の集中化、など柔軟に軌道修正を行うとともに、積極的に3グループ間で協力関係を築いてほしい。

具体的には以下のとおりである。

- ① バイオインフォマティクス技術については理論的な成果は積み上げられているが、検証実験については不十分であり、プロジェクト全体に対する貢献度を上げるためにも、開発した予測技術の有効性を明確にすべきである。
- ② 支援技術・ツールの開発においては各テーマの有機的な連絡、研究の必然性が不明である。従来技術と対比した上での活用法、残された解析課題などを明確にして欲しい。
- ③ 機能性RNAの機能解明については互いに独立な多数の研究グループの集まりのような印象がある。プロジェクト後半は、良い芽を出している研究をいくつか選択して、実用化・事業化を目指した研究を行うために、テーマの推進・連携方法について工夫すべきである。

### 5. 2-3 対応を要する指摘事項への対応

#### 指摘事項

① バイオインフォマティクス技術については理論的な成果は積み上げられているが、検証実験については不十分であり、プロジェクト全体に対する貢献度を上げるためにも、開発した予測技術の有効性を明確にすべきである。

#### 【対応】

バイオインフォマティクス技術については、機能解析グループとも連携しつつ、検証実験を計画し、機能解析候補の絞り込みを通じてインフォマティクス技術の有効性を示す取り組みを進めた。

#### 指摘事項

② 支援技術・ツールの開発においては各テーマの有機的な連絡、研究の必然性が不明である。従来技術と対比した上での活用法、残された解析課題などを明確にして欲しい。

#### 【対応】

支援技術・ツールの開発においては従来技術と比較した優位性及び問題点の明確化、開発技術の活用法の検討による必要性の明確化を行い、テーマの絞り込みと予算の集中化を行った。

指摘事項

③機能性RNAの機能解明については互いに独立な多数の研究グループの集まりのような印象がある。プロジェクト後半は、良い芽を出している研究をいくつか選択して、実用化・事業化を目指した研究を行うために、テーマの推進・連携方法について工夫すべきである。

【対応】

機能性RNAの機能解明については各グループの有するリソースや情報の共有などを通じて連携を進めるとともに、産業競争力に寄与する先進的な基盤技術になるテーマ、競合優位性を獲得しうるテーマの選択、強化を行い、研究の促進を図った。



### Ⅲ. 研究開発成果について

#### 1. 事業全体の成果

##### 1.1 成果概観

2005年7月から2010年2月まで、約5年弱の期間を費やして、「機能性 RNA プロジェクト」が産官学連携の下で実施された。ちょうど今世紀に入る頃に発見された種々の形態をもつノンコーディング機能性 RNA を検出・単離するツールを開発し、その機能を解明するという本プロジェクトは、時代の流れに即したものであった。それは単に生命現象における機能性 RNA の機能を探求するにとどまらず、多くは未知の機能の解明を通じてガンやもろもろの疾病に対する医薬品の開発や再生医療に資する基盤技術を確立するという極めて発展性が高く、応用範囲の広いテーマである。

世界的な激しい競争下において独自性と優位性を確保するために設定された、①機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発、②機能性 RNA 解析のための支援技術・ツールの開発、③機能性 RNA の機能解析という3研究開発項目が、それぞれの目標の実現に向けて努力するとともに、3者が相互に緊密な連携を取りながら研究を推進することにより、数々の重要で顕著な成果を挙げる事ができた。

○機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発では、機能性 RNA の網羅的予測を可能にし、世界最高レベルの2次構造を考慮した RNA 配列情報技術を数多く開発することに成功し、今後 RNA 創薬のための有用な解析ツールとなる予測機能性 RNA のカスタムマイクロアレイ（連携研究による）や、機能性 RNA だけでなくゲノム中のさまざまな因子を網羅的に解析するための情報基盤として有益な機能性 RNA データベースを開発するなど、多くの実用化に近い成果を挙げた。

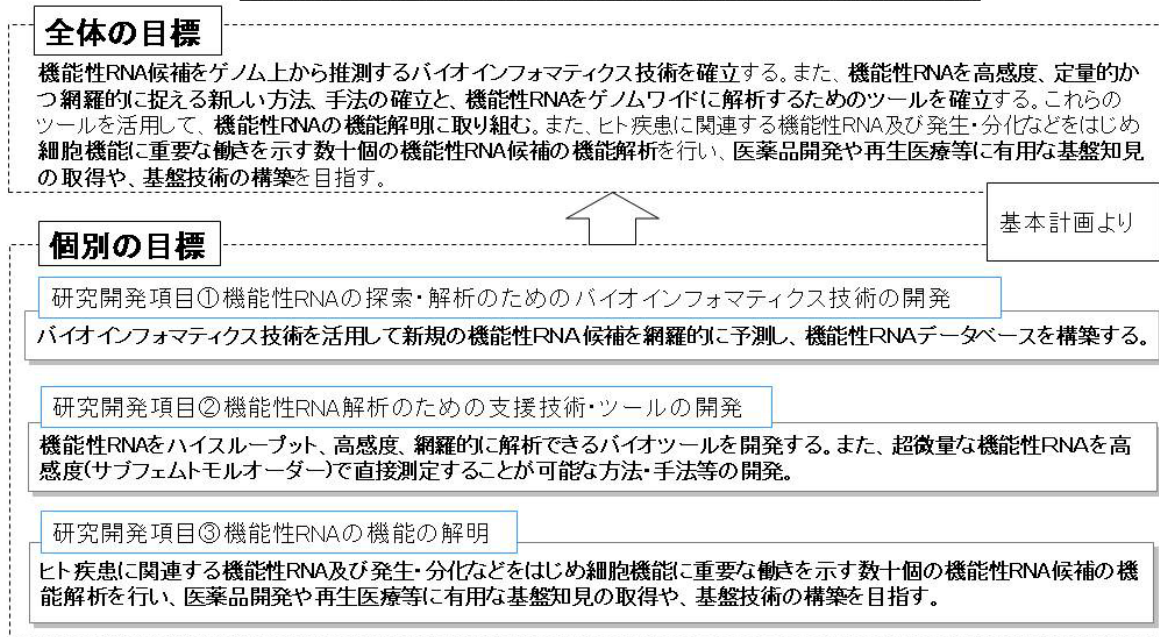
○機能性 RNA 解析のための支援技術・ツールの開発では、世界標準を目指した核酸医薬の薬物動態の解析に役立つ RNA-MS と情報解析技術、RNA の全自動合成精製技術など実用化に近い成果や、製品化に向け検討中の miRNA 用高感度マイクロアレイやすでに研究用試薬として販売を開始した高品質で安価な RNA 化学合成法など、もっとも出口に近い成果を挙げている。

○機能性 RNA の機能解析では miRNA 機能を利用した再生医療技術、アレルギー疾患治療技術、ガン治療技術、あるいは iPS 細胞の効率良い作出技術などの他、診断マーカーとしての長鎖 ncRNA やアンチセンス RNA の利用技術の基盤を確立した。

これらの成果は機能性 RNA 研究を飛躍的に発展させ、本分野における我が国の優位性を確立することに大きく貢献した。さらに、特に機能解析では再生医療、難疾患治療等医療・診断分野における新産業の創出などに貢献し得る成果を挙げた。miRNA の機能解析では世界的な激しい競争下で評価すべき上記の成果を挙げたが、長鎖 ncRNA の機能解析は、miRNA の研究と比較すれば世界的に見てもやっと研究の端緒に辿り着いたばかりの分野であり、まだ解決すべき問題は山積しており、今後の飛躍的な発展が期待される。

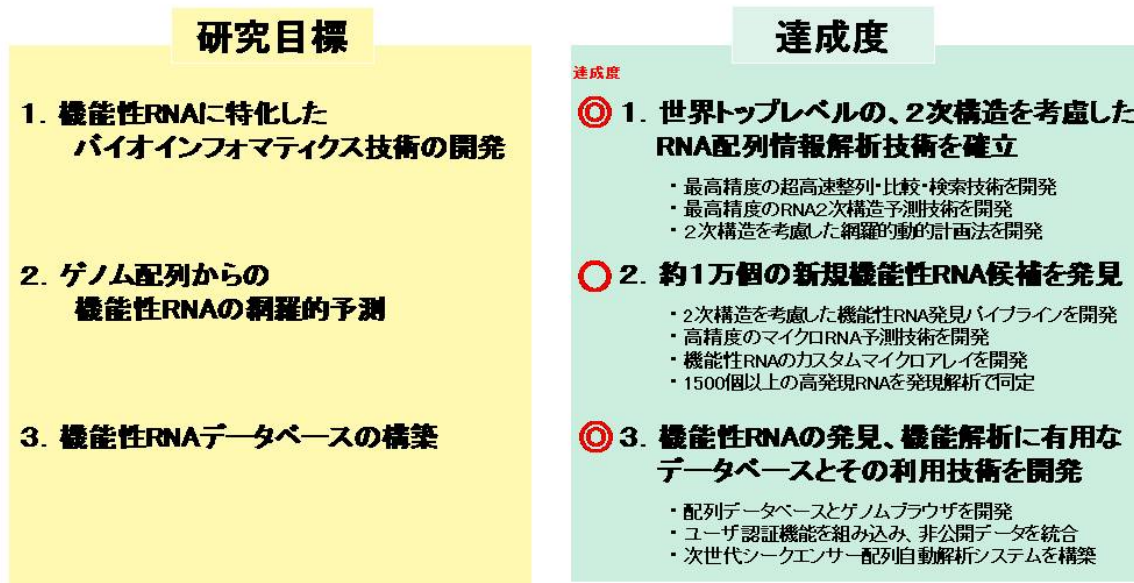
## 1.2 最終目標と達成度

### 研究開発項目ごとの個別目標を達成し全体目標に至る。



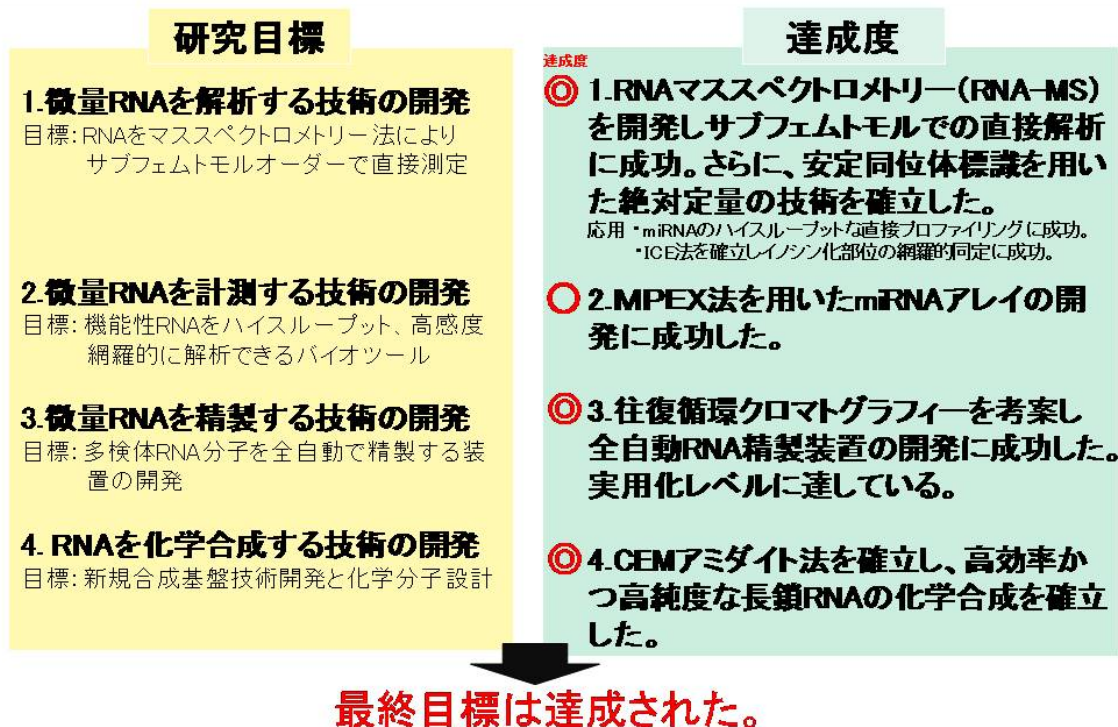
26/53

## 研究開発項目① 目標達成まとめ

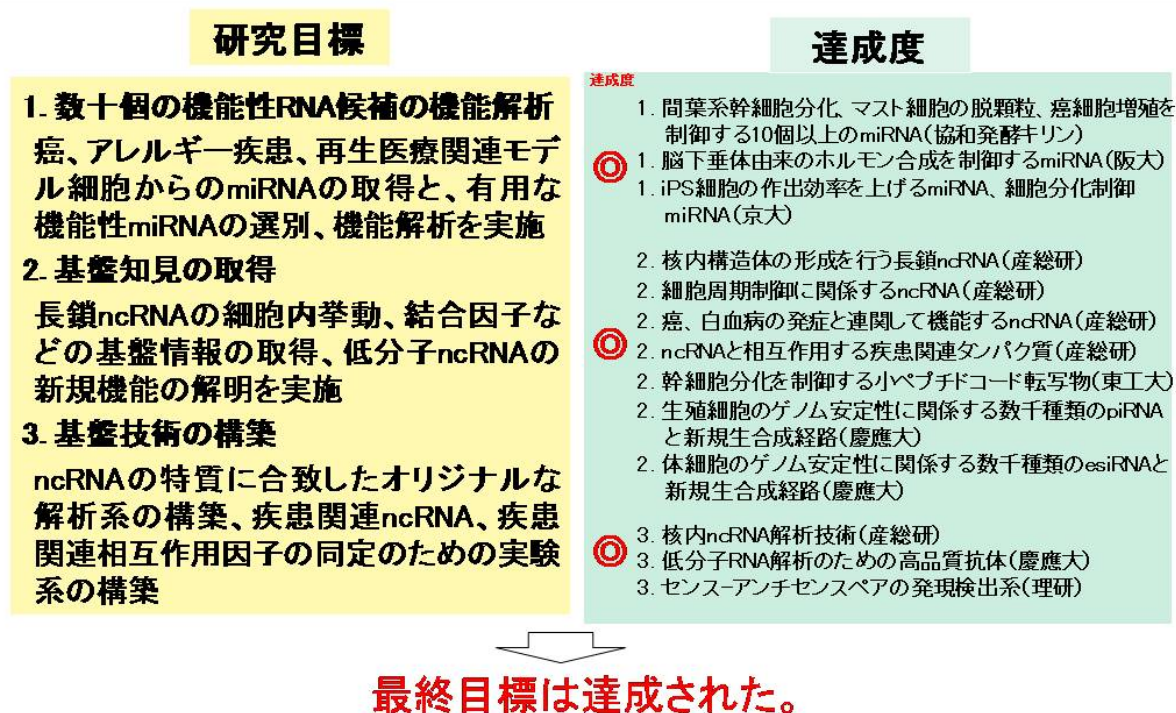


最終目標は達成された。

## 研究開発項目② 目標達成まとめ



## 研究開発項目③ 目標達成まとめ



### 1.3 達成度各論

#### 1.3.1 ① 機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発

##### 【最終目標】

バイオインフォマティクス技術を活用して新規の機能性 RNA 候補を網羅的に予測し、機能性 RNA データベースを構築する。

##### 【目標に対する達成度】

① RNA の発見、機能解析に重要な、2 次構造を考慮した配列情報解析技術に関して、画期的な技術を開発に成功し、数々の有用なソフトウェアを開発した。理論面においても、期待精度最大化原理を発展させた新理論を展開し、2 次構造予測、構造アラインメントの精度や速度では世界最高レベルを達成した。また、高度な動的計画法を駆使し、RNA の相互作用、塩基置換によるエネルギー変化などの解析に必須なソフトウェアを開発した。また、RNA 配列の比較・検索に有効なカーネルを設計し、snoRNA の発見に有効であることを実証した。さらに、開発したソフトウェアは①-2 機能性 RNA 発見に提供し、Nature、PNAS 等の一流誌（4 報）にて掲載された共同研究成果に結びついた。

このように RNA 配列情報解析技術（バイオインフォマティクス技術）については、世界最高レベルの技術を数多く開発することに成功し、目標を大きく上回る成果である。

② 機能性 RNA を効率よく発見するための基盤技術の開発に成功し、ヒトゲノムから 10,000 個以上の機能性 RNA をバイオインフォマティクスにより予測し、カスタムアレイによる解析によって、1,500 個以上の高発現機能性 RNA を発見した。また、進化的に保存したマイクロ RNA に対する最高精度の予測技術の開発に成功すると共に、mRNA の構造制約を評価して標的 mRNA を予測する技術を開発した。また、偽遺伝子の特徴を利用した機能性 RNA 予測手法を開発し、癌マーカー遺伝子や新規 snoRNA を発見した。

このように新規の機能性 RNA 候補の網羅的予測については、目標を達成している。

③ 50 万配列以上を網羅した機能性 RNA データベースを構築し、公開した。配列データベースである fRNAdb と、機能性 RNA に特化したゲノムブラウザである UCSC Genome Browser for functional RNAs を連携させて開発し、既知の機能性 RNA 情報、予測機能性 RNA 情報、miRNA 標的など数多くの有用な情報を DB 上で提供している。また、ユーザ認証機能を組み込み、共同研究に活用するなど、プロジェクト内の情報基盤を提供した。さらに、次世代シーケンサー配列自動アノテーションシステムを開発した。これらのデータベース、システムは、エピゲノム解析など今後様々なプロジェクトに活用できる情報基盤である。

このように機能性 RNA データベースの構築については、実用的で網羅的なデータベースの開発に施工したことに加え、ユーザ認証機能の付加とその活用におけるプロジェクト内共同研究の成功は、目標を上回る成果であった。

以上を総合し、本研究開発項目は目標を大きく上回る成果を上げたと考えられる。

### 1.3.2 ②機能性 RNA 解析のための支援技術・ツールの開発

#### 【最終目標】

機能性 RNA を高感度、定量的かつ網羅的に捉える新しい方法、手法の確立と、機能性 RNA をゲノムワイドに解析するためのツールを確立する。これらのツールを活用して、機能性 RNA の機能解明に取り組む。

#### 【目標に対する達成度】

①RNA を高感度で直接解析するための手法として、RNA マススペクトロメトリー (RNA-MS) を開発し、中間年度ですでにプロジェクトの最終目標値であるサブフェムトモルオーダーでの RNA の微量解析に成功した。本手法を駆使することで、piRNA の末端修飾構造の同定、miRNA の選択的安定化機構の解明、miRNA 直接プロファイリング、新規 RNA 修飾の発見など、これまでの手法では不可能であった様々な解析に成功した。また、膨大な質量情報を情報処理するアルゴリズムやソフトウェアの開発にも成功し、知財戦略的にも優位に立った。さらに RNA-MS を支援するツールとして、往復循環クロマトグラフィー法を考案し、微量 RNA の全自動精製装置の開発にも成功した。RNA-MS は、cDNA 解析に頼らない全く新しい RNA の解析技術であり、機能性 RNA 研究のみならず、RNA 創薬や核酸医薬の開発を強力にサポートする基盤技術に成長した。本開発テーマは、最終目標値をクリアすることにより、加速財源の投入を得ることが出来、多くの研究業績を残すことができた。

②RNA の網羅的解析技術としては、ヒトの RNA に大量に含まれるイノシン化修飾を特定する手法の開発を行った。ヒト脳由来 RNA には大量のイノシン化修飾が存在することが知られ、高次なレベルでの脳機能に関わると考えられている。イノシン化部位を網羅的に同定するために、独自に開発した ICE 法と、次世代シーケンサーを組み合わせた手法 (ICE-Seq 法) を開発し、ヒト脳のトランスクリプトームから、現在までに 3 万箇所を超えるイノシン化部位を同定した。このうちの約 80% の部位は計算科学では予測不能な新規部位であった。これらの多くは miRNA による翻訳抑制を調節する役割があることが示唆された。今後は脳神経疾患と RNA エディティングの変動解析を行っていく予定である。

③RNA のハイスループット、高感度検出に関しては、多重伸長反応 (MPEX) 法や光化学反応法を利用した新規の手法を開発した。MPEX 法に関しては miRNA をプロファイリングするためのマイクロアレイを開発し商品化へのめどをつけた。

④RNA の化学合成技術も実用化に成功し高品質かつ安価な核酸医薬の供給が可能になり、実用化と事業化に成功している。本手法により、長鎖 RNA の化学合成に成功し、生物活性を確認することができた。高品質かつ安価な合成 RNA を供給するシステムが整い、今後は

これらの技術を診断分野や核酸医薬の開発に積極的に活用する予定である。

以上より最終目標に対する達成度は100%と言える。

### 1.3.3 ③ 機能性 RNA の機能解析

#### 【最終目標】

ヒト疾患に関連する機能性 RNA 及び発生・分化などをはじめ細胞機能に重要な働きを示す数十個の機能性 RNA 候補の機能解析を行い、医薬品開発や再生医療等に有用な基盤知見の取得や、基盤技術の構築を目指す。

#### 【目標に対する達成度】

①疾患や再生医療のための重要な新技術となりうる有用な機能性 miRNA が数多く同定された。協和発酵キリンググループが同定したマスト細胞からの脱顆粒効率を制御する miRNA は、今後のアレルギー疾患の治療技術につながる重要なものである。同グループによる癌細胞の増殖を制御する miRNA についても、同様の活性を有する既知の miRNA よりも高効率で機能するものも取得されており有望である。上記の2つの現象については、ポジティブな効果を果たす miRNA とネガティブな効果を果たす miRNA がそれぞれ取得されていることも特筆に値する。つまりポジティブ因子として miRNA に対しては、その機能を抑制するための核酸誘導体を投与する戦略を立てることができ、一方でネガティブ因子として miRNA に対しては、miRNA そのものを核酸医薬品として用いて細胞に投与する戦略を立てることが可能になる。同グループは、miRNA を投与した培養細胞をマウスに移植しゼノグラフト化したものを長期に追跡し、miRNA 効果が長期に渡り持続していることまで確認している。こうした創薬開発のための基盤的検定を積み重ねることによって、miRNA を用いた新しい医薬品の開発に結びつく事が期待される。

②京都大学グループでは、iPS 細胞の作出効率を著しく増大させる miRNA を新しく同定することに成功した。iPS 細胞の作出は、効率だけでなく、高品質な iPS 細胞を作り出すことなど、今後益々技術向上が求められることが推測されるが、miRNA はそうした過程で重要な因子としてクローズアップされることが期待される。本プロジェクトの成果は、その先駆けとなるもと位置づける事ができ、今後の iPS 細胞作出技術向上への貢献と、作出メカニズム解明の糸口として利用されることを期待したい。

③ゲノムの暗黒物質とも呼ばれ機能がほとんど未知であった長鎖 ncRNA についての様々な新しい特徴が明らかになった。産総研グループは、長鎖 ncRNA の特徴として細胞核内で機能すること、特定の組織で特異的に機能するものが多数を占めることを明らかにした。細胞核内の RNA を機能解析するための新しい解析ツールとして、アンチセンス核酸を用いた核内 RNA ノックダウン法を開発し、この方法を駆使した核内 ncRNA の新しい機能を解明した。特に細胞内構造体の構築を司る ncRNA 機能や、細胞周期特異的な DNA 複製時の制御を

行う ncRNA 機能、クロマチン上での発現制御機能などの発見は、特筆すべき成果といえる。一方で、ncRNA が相互作用タンパク質と共に機能体を形成することに注目し、微量の ncRNA と相互作用する因子の新しい同定法を考案し、新しいタンパク質と ncRNA の相互作用を明らかにした。こうした研究の進め方は、今後の長鎖 ncRNA の機能解析のための先駆けとなる研究手法となることが期待される。また相互作用因子の解析や詳細な発現解析から長鎖 ncRNA や理化学研究所グループの解析によって得られた多くのアンチセンス RNA が癌、神経変性疾患などの難治性疾患と関わる可能性が次々と明らかになってきた。今後の詳細な機能解析を継続することによって ncRNA と疾患の接点が明らかになると期待できる。

④慶應大学グループによって、miRNA に続く新しい低分子 ncRNA の機能と相互作用因子が発見された。低分子 ncRNA が共通して結合している argonaute ファミリータンパク質を特異的な抗体を用いて解析したことによって、生殖細胞特異的な piRNA 群や体細胞由来の esiRNA 群といった新たなカテゴリーに属する低分子 ncRNA が発見された。これらの RNA は、miRNA 機能とは異なり、ゲノムを乱す様々な可動性因子の働きを抑える役割を果たしていることが推測されている。こうして数ある Argonaute タンパク質と低分子 ncRNA との的確な組み合わせによって機能分化がなされて、様々な様式で細胞内の重要な働きを行っていることを世界に先駆けて発見した点は特筆に値するといえる。

①で得られた成果は、医薬開発に向けた次のステップの研究に進行中である。次なるステップには、これらの核酸をデリバリーするための技術革新が求められている。一方、③で得られた ncRNA に関する基盤知見は、これまで配列情報だけで類推されてた ncRNA 機能に、新しい機能概念を加えたこと、さらには、ncRNA のための機能解析系が整備されたことが大きな成果であると言える。長鎖 ncRNA に関しては、プロジェクト期間中に世界的にも大きな進展が見られなかったことから、本プロジェクトで整備した研究体制を継続することによって、今後重要な発見に至る事が期待される。また既にこのプロジェクト中にも既に一端が見出されたように、基盤的な知見には必ず疾患との接点が見出され、こうした接点を今後詳細に解析することによって全く新しい医薬品開発のコンセプトが生まれることが期待される。そのような ncRNA ならではの独自機能の利用こそが真の意味でのオリジナルな RNA 医薬品となるものと考えている。

以上のように、プロジェクト開始時に想定した成果を大きく上回るオリジナリティの高い成果が数多く得られ、最終目標に対する達成度は100%と言える。

## 2. 研究開発項目毎の成果

本プロジェクトは、以下に示す3つの研究開発項目から構成されているため、3つの研究開発項目毎に、成果を纏め、既に設定されてある最終目標に対する達成度との観点からの記述と各研究室での研究成果の要約に分けて纏めた。以下にそれらを記載する。

- ① 機能性RNAの探索・解析のためのバイオインフォマティクス技術の開発
- ② 機能性RNA解析のための支援技術・ツールの開発
- ③ 機能性RNAの機能解析

上記の研究を達成するために、産官学の共同研究体制を産業技術総合研究所内に組み、さらにこれを強化するために外部の研究機関の連携のもと研究開発を行った。



## 2. 1 機能性 RNA の探索・解析のためのバイオインフォマティクス技術の開発

### 【成果要約】

本研究開発項目では、ゲノム配列・cDNA 配列などからの情報技術を用いて機能性 RNA を網羅的に発見し、機能性 RNA データベースを構築し、その機能を予測して実験による計測や機能解析にも貢献する。情報技術による機能性 RNA の発見技術は未成熟なため、機能性 RNA に特化した配列情報解析技術の開発と、新旧技術の組み合わせによる機能性 RNA の網羅的发现を平行して実施し、既知の機能性 RNA、網羅的な機能性 RNA 発見の結果およびその他の情報解析結果を統合した機能性 RNA データベースを構築する。中間目標（平成 19 年度）は、1）ゲノム配列から機能性 RNA を網羅的に検出するバイオインフォマティクス技術を確立する、2）機能性 RNA の機能を予測するための情報技術を開発する、の 2 点であり、最終目標（平成 21 年度）は、バイオインフォマティクス技術を活用して新規の機能性 RNA 候補を網羅的に予測し、機能性 RNA データベースを構築することである。

機能性 RNA の発見のためには、配列の類似性だけでなく、その二次構造に基づいた解析を行わなければならない。①-1 「機能性 RNA に特化したバイオインフォマティクス技術の開発」では、膨大な計算時間と記憶容量が必要であった従来手法の限界を克服するため、さまざまなアルゴリズム・ソフトウェアの開発を行った。その結果、二次構造を考慮した RNA 配列の構造アラインメントを高精度、超高速に行うソフトウェア、1 本の RNA 配列の 2 次構造および RNA 配列群の共通 2 次構造の予測で最高精度のソフトウェア、2 次構造と配列の両方の類似性をカーネル法によって取り込んで既知ファミリーの機能性 RNA の検索を高精度に行うソフトウェア等の開発に成功した。また、共通二次構造・二次構造モチーフを抽出するソフトウェア、RNA のアクセサビリティなど局所的なエネルギー構造を計算するソフトウェア、塩基置換に伴うエネルギー、構造等の変化を配列のすべての位置について網羅的に計算することのできる動的計画法に基づくソフトウェアなど、ゲノムレベルの網羅的な機能性 RNA の発見と機能予測や、miRNA、siRNA の標的予測、核酸医薬への応用に不可欠な技術を数多く開発した。

①-2 「ゲノム配列からの機能性 RNA の網羅的予測」では、開発したアルゴリズム・ソフトウェアを積極的に活用し、ゲノムのシンテニー領域、ヒトゲノム内の相同領域等を中心に解析、約 20 万箇所の候補領域を抽出し、その中から 1 万箇所以上の有望な機能性 RNA 候補を抽出した。また、進化的に保存された miRNA の予測を高精度に行う手法を開発し、330 個の新規 miRNA 候補を発見した。これらの新規機能性 RNA 候補についてカスタムマイクロアレイを設計して発現解析を行い、1500 個以上の高発現の機能性 RNA を発見した。RNA の分解に関与する Exosome のノックダウン実験による発現解析を機能解析グループと共同で行い、発現量が必ずしも高くない機能性 RNA を効率よく発見出来る手法を構築した。我々が行った機能性 RNA 候補の抽出、すなわちインフォマティクス技術とマイクロアレイを組み合わせたパイプラインは、発現している機能性 RNA を効率よく発見するために有効な基

盤技術であると考えられる。

①-3「機能性 RNA データベースの構築」では、既知の機能性 RNA に関する配列、文献、マッピング情報と、網羅的に予測された機能性 RNA 候補の情報を統合することにより、機能性 RNA の発見・機能予測を支援する情報基盤を提供することを目的に、UCSC ゲノムブラウザに独自の解析結果・情報を統合した UCSC GenomeBrowser for Functional RNA と、機能性 RNA の配列情報ブラウザ fRNAdb を開発した。さらに、各ユーザに固有の非公開データと公開データを同時に表示して解析を行ことのできるユーザ認証を組み込み、プロジェクト内の共同研究を促進する情報基盤を提供した。機能性 RNA データベースを活用した機能解析グループとの共同研究により重要な新規機能性 RNA を発見し、トップジャーナルに論文発表した。また、次世代シーケンサーによる発現解析から得られる大量の配列データを自動的に解析し、その帰属の決定とゲノムへのマッピングを行うシステム GIGANT を開発した。

## 【研究開発の方針】

本研究開発項目の目的は、機能性 RNA 配列を予測し、その機能を解明するためのバイオインフォマティクス技術を開発することである。機能性 RNA を情報技術によって予測し、予測した機能性 RNA 候補の機能解析を実験グループと連携して行い、その成果を蓄積すると共に解析結果を新たな発見に結び付けるため、本研究開発項目では、基本計画に示された3つの課題、①-1「機能性 RNA に特化したバイオインフォマティクス技術の開発」、①-2「ゲノム配列からの機能性の網羅的予測」、①-3「機能性 RNA データベースの構築」を行った。

mRNA においては、そのコードするタンパク質の種類にかかわらず、3文字ずつアミノ酸に翻訳されるための明確な情報構造を持っているが、個々の機能性 RNA ファミリーはそれぞれ全く異なった配列・構造上の特色を持っているから、機能性 RNA を情報技術によって発見する一般的な手法を構築することには無理がある。しかしながら、既知の機能性 RNA ファミリーに関する知見から、同一の機能を持つと考えられる機能性 RNA ファミリー内の RNA 同士は、類似の2次構造をもつことが期待されるから、2次構造を考慮した配列の類似性、保存性に着目した解析によって新規機能性 RNA を発見しようとするのは自然な考えである。

既に発見手法が確立している tRNA、rRNA に加え、miRNA、snoRNA のような既知の機能性 RNA ファミリーの新規 RNA 遺伝子を発見する場合には、それぞれの RNA ファミリーに特有な2次構造・配列の特徴をもとにゲノム上を検索することが有効であることは、tRNA-scan、Infernal 等のツールの有効性から明らかであった。ところが、代表的な既知機能性 RNA ファミリーである miRNA、snoRNA の発見でさえ、十分な精度で行うことのできるツールは存在しなかった。また、miRNA、snoRNA の標的遺伝子(RNA)の予測に関しても、配列相補性のみに着目した既存手法では信頼できる結果が得られない状況であった。

タンパク質コード遺伝子(mRNA)の予測においてゲノム間比較での保存領域に着目すると同様に、新規機能性 RNA ファミリーに属する機能性 RNA 予測においても、ゲノム間・ゲノム内で保存された領域に着目して予測を行うことは有効であると考えられる。実際、QRNA、RNAz などのツールが本プロジェクト開始時において既に開発されていた。しかしながら、2次構造の保存性の厳密な検出は計算複雑度が高く(2本の RNA 配列に対する計算時間で配列長の6乗、記憶容量で配列長の4乗に比例)、ゲノムワイドな大規模解析が不可能だったこともあり、保存領域の抽出やその解析技術では2次構造を無視した解析の精度が低く、予測結果の信頼性にも問題があった。さらに、さらに、ヒトゲノムの大部分が実際には何らかの形で RNA に転写されていることが次第に明らかになるに及んで、新規機能性 RNA 候補を情報技術によって予測し、その候補配列が RNA として発現していることを実証するだけでは、有用な発見を行ったとはみなすことができない状況となってきた。このため本研究開発項目では、単に2次構造が保存されて転写されていることを実証するだけでなく、組織特異的に発現する候補野絞り込みなど、機能解析と連携した予測・解析を行

うこととした。

以上から明らかなように、2次構造を考慮してRNA配列を比較、整列、モデル化、検索する情報技術は、機能性RNAの発見とその機能解析を行う上でカギとなる技術であるにもかかわらず、タンパク質アミノ酸配列やゲノムDNA解析における標準ツール（BLASTなど）は存在せず、新たな配列情報解析技術の開発抜きには、新規機能性RNAの発見は難しい情勢であった。そのため本研究開発項目では、RNA配列の高速高精度な構造アラインメント（2次構造を考慮した配列アラインメント）手法、2次構造と配列の両方を考慮したRNA配列の比較・検索、RNAの2次構造予測、2次構造を考慮したRNA相互作用の予測等のRNA配列情報解析技術の開発と、それらの技術を組み合わせた新規機能性RNA予測を連携しながら同時に進めることとした。また、予測した約1万個の機能性RNA候補についてカスタマイクローアレイを設計し、様々な組織での発現を解析して有用な機能性RNA候補を絞り込み、機能解析グループと連携した研究開発を行った。2次構造を考慮したRNA配列情報解析技術の未発達と並んで、機能性RNAの発見、解析の進展を阻んでいたのがデータベースの欠如である。ほとんどのゲノム解析プロジェクトやゲノムデータベースが、タンパク質コード遺伝子を中心に構築されてきたために、機能性RNAに特化し、かつ網羅的なデータベースは不十分であった。たとえば、本プロジェクト開始時点では、機能性RNAに特化した網羅的データベースであるRfamには、ゲノム配列への転写情報はなく、逆にUCSCゲノムブラウザにはRfam配列のアノテーションはなかったのである。本研究開発項目では、単に既知の機能性RNA配列のデータを蓄積するだけでなく、予測した機能性RNA配列、予測の元となった様々な情報解析結果、次世代シーケンサーによって読まれた機能性RNA断片のデータなどを集積し、新規機能性RNAの発見と絞り込み、機能解析に活用できるデータベースの開発を行うことを目標に、研究開発を進めることにした。同時に、実験グループから提供されるRNA配列データを解析し、開発したデータベースを活用して新規機能性RNAの発見する研究にも取り組んだ。

**【各研究室における成果】**

**2. 1. 1. 機能性 RNA に特化したバイオインフォマティクス技術の開発**

**JBIC 集中研①**

共同実施先： 産総研①、東京大学（1；浅井）、慶応義塾大学（1；榊原）

みずほ情報総研

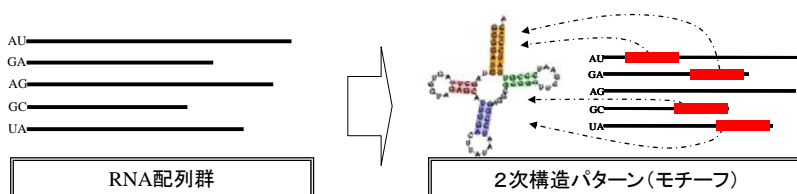
**2. 1. 1. 1. 機能性RNAのための総合的な配列解析技術の開発**

機能性 RNA をバイオインフォマティクスにより解析するためには、従来の配列解析技術だけでは不十分である。本研究開発項目においては、RNA の 2 次構造予測、モチーフ探索、アラインメントなど、ほとんどの情報解析の基盤となるバイオインフォマティクス技術の開発を行った。

## 2. 1. 1. 1. 1. RNAmine (RNA 配列群からの頻出する 2 次構造パターン抽出ソフトウェア)

### (1) 背景

機能性 RNA の多くは、機能と密接に関連した特長的な 2 次構造を有することが知られている。従って、RNA 配列群に頻出する 2 次構造パターンを同定することは、機能性 RNA の機能と関連する可能性が高い 2 次構造を同定することにつながる。この際、配列からの 2 次構造予測の精度には限界があるため、決められた 2 次構造を利用するのではなく、配列の潜在的な 2 次構造情報を利用した 2 次構造パターンの同定が望まれる (図 1)。



- **入力:** RNA 配列群
  - アラインメントされている **必要なし**
  - 2次構造がわかっている **必要なし**
- **出力:** 入力配列群中に「高頻度で」出現する **2次構造パターン(モチーフ)**
  - 複数存在する場合は **すべて** 導出する

図 1. RNAmine の概要

### (2) 手法

以下の 2 つの手順により、与えられた RNA 配列群に頻出する 2 次構造パターンを同定する。

#### (i) RNA 配列のグラフによる表現 (ステムパターン) とラベルタクソノミ

RNA 配列をステムグラフと呼ばれるラベル付き有向グラフにより表現する。ステムグラフの各頂点は RNA 配列のステム候補、ラベルはステム集合をその類似性を用いてクラスタリングすることにより付与する。また、有向辺およびそのラベルは 2 つのステム間の関係に応じて与える。このようにして得られるステムグラフは RNA の潜在的な部分 2 次構造を可能な限り考慮した表現となる。さらに、本研究では、ノードのラベルを一つに固定するのではなく、ラベルタクソノミと呼ばれる階層的なデンドログラムとして表現した。これにより、非常に柔軟なラベルのマッチが可能となった。

#### (ii) コストがある閾値以下の頻出クリークパターンの抽出

上述の表現方法により、配列群から頻出する 2 次構造パターンを抽出するという問題は、ステムグラフ集合からコストがある閾値以下のクリークパターンを抽出する問題として定式化される。これは、データマイニングの業界で近年活発に研究がなされているグラフマイニング問題と考えることができる。我々は、独自に、グラフマイニン

グアルゴリズムを考案した。

### (3) 結果

RNAmine を利用して、各 RNA 配列の 2 次構造を利用した際の 2 次構造予測結果を表 1 に示す。

表 1. RNAmine の性能評価

Family	RFAM_ID	#seqs	length	%id	RNAmine			CMfinder	comRNA	RNAalifold	RNAfold	RNAsubopt	str/seq
					1	5	10	MCC	MCC	MCC	MCC	MCC	
Cobalamin	RF00174	50	203.2	43	0.41	0.52	0.53	0.54	0.00	0.47	0.34	0.44	119.6
Lysine	RF00168	50	181.6	46	0.80	0.85	0.86	0.79	0.21	0.35	0.64	0.74	112.3
Purine	RF00167	37	99.6	53	0.83	0.90	0.91	0.89	0.00	0.52	0.73	0.81	8.3
RFN	RF00050	48	137.2	64	0.62	0.71	0.74	0.41	0.00	0.57	0.44	0.52	29.4
S_box	RF00162	50	110.4	61	0.77	0.82	0.84	0.78	0.29	0.48	0.64	0.76	35.1
Tymo_tRNA-like	RF00233	27	82.6	66	0.76	0.88	0.88	0.93	0.55	0.51	0.60	0.72	10.3
glmS	RF00234	14	177.6	55	0.80	0.86	0.90	0.88	0.47	0.35	0.58	0.66	30.7
tRNA	RF00005	50	73.4	40	0.75	0.84	0.84	0.78	0.00	0.37	0.60	0.73	8.5
				average	0.72	0.80	0.81	0.75	0.19	0.45	0.57	0.67	44.3

RFAM\_ID: ID number in Rfam database (<http://www.sanger.ac.uk/Software/Rfam/>). #seq: the number of sequences in each family. length: average length of sequence in each family. %id: average sequence identity calculated by alifat program. MCC: average MCC among sequences. Best MCC among top 1, 5 and 10 structures are shown in result of RNAmine. For comRNA and RNAsubopt, the best MCC among predicted common secondary structures is shown (if comRNA produced no motif, MCC is 0 in this table). str/seq (for RNAsubopt): the average number of predicted suboptimal secondary structures per sequence. The definition of MCC is found in the supplementary paper.

### (4) 成果の活用と公開

プロジェクトでは、RNAmine を用いて実際に機能性 RNA 候補の同定を行った。その方法および結果については後述する。

RNAmine は Hamada et al., *Bioinformatics* 22(20), pp. 2480–2487, 2006 で論文発表し、そのアルゴリズムは特許出願した。

<http://software.ncrna.org/cgi-bin/index.cgi?page=Input&program=rnamine> から、WEB サーバが利用可能である (図 2)。

<http://www.ncrna.org/>

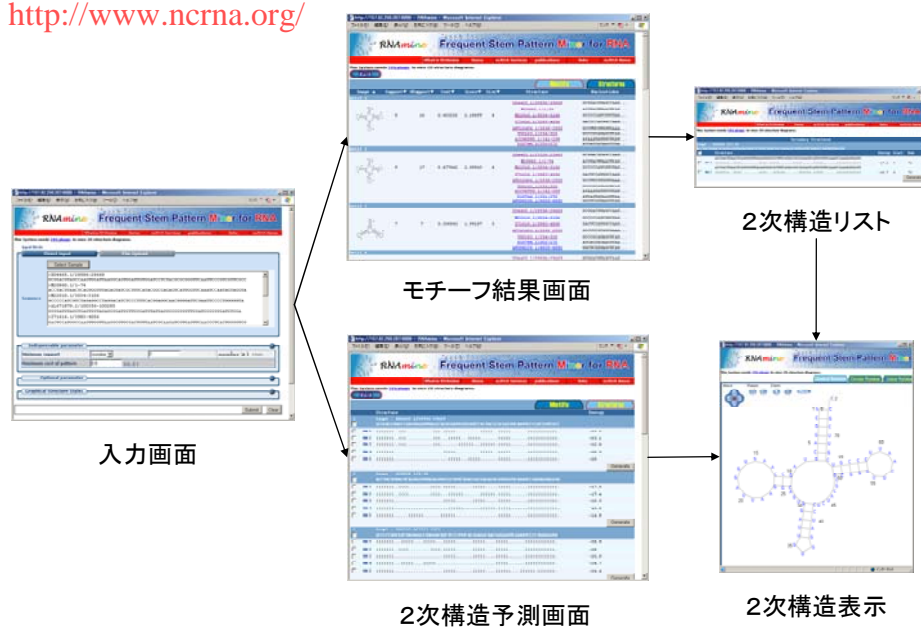


図2. RNAmine の Web インターフェース



## 2. 1. 1. 1. 2. RNAclique (RNA 配列群から類似した局所 2 次構造クラスタ探索ソフトウェア)

### (1) 背景

前述の RNAmine はグラフマイニング手法を用いるため、多くの計算量を必要とする。そのため、実際の解析においては前処理を行い、候補配列群を絞る必要があった。しかしながら、実験や計算機によって予測された多数の機能性 RNA 候補に加え、ゲノム中には HAR や TFR など多数の機能未知の興味深い領域が存在する。これらの領域を解析の対象とするためには、大規模な配列群から 2 次構造の類似した 2 次構造のクラスタを抽出する手法が必要である。

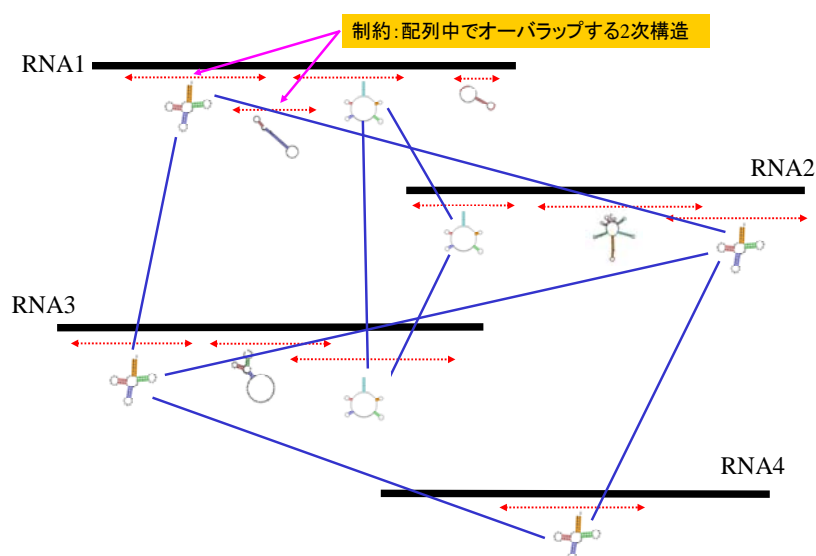


図 3. RNAclique の概要

### (2) 手法

RNAmine とは異なるアプローチを用いて、大規模な配列群から類似した局所 2 次構造を抽出する手法 RNAclique の開発を行った。RNAclique は以下の手順に従い 2 次構造の類似したクラスタを抽出する (図 3)。入力の RNA 配列群  $\{S(i) : i=1, 2, \dots, N\}$  に対して、以下のステップで処理を行う。

#### (i) 局所安定 2 次構造の候補集合の計算

各 RNA 配列に対して、局所安定 2 次構造の候補集合を計算する。今回は RNALfold と呼ばれる局所安定 2 次構造予測ソフトウェアを用いた。

#### (ii) 配列ベースのフィルタリング

得られた局所安定 2 次構造を有する部分配列群に対して、比較的ゆるいパラメータを用いて all-to-all で wu-blast 検索を行い、次のステップで 2 次構造の類似度を計算するペア候補をフィルタリングする。

#### (iii) 2 次構造の類似性計算

上述のフィルタリングを通過した局所2次構造のペアに対して、2次構造の類似性を計算する。今回はソフトウェア RNAforester (with RIBOSUM80-85 行列) を用いて2次構造の類似性を計算した。

(iv) 制約付きグラフの構築

局所2次構造をノード、類似性がある閾値以上の局所2次構造間に辺を結ぶことによりグラフを構築する。また、同じRNA配列に含まれるオーバーラップしている局所2次構造は、同じクラスタに含まれては困るので、同じクラスタに入ることの出来ないノードのペアを制約として与える。このため、ここで構築されるグラフを制約付きグラフと呼ぶ。

(v) 制約付き偽クリーク探索

上記で構築した制約付きグラフから、制約を満たす偽クリーク部分グラフを抽出する。一般にこの問題はNP困難な問題であるため、本研究ではGRASPと呼ばれるヒューリスティックアルゴリズムを適用した。

(vi) 得られたクラスタ集合の事後処理

得られたクラスタ集合に対して、ランキング付けを行う。

(3) 計算機実験

計算機実験の結果、提案手法は他の手法に比べてクラスタを抽出する精度が優れていることがわかった。

**Modified method 1:** 局所2次構造をオーバーラップしないように(エネルギーが低い順)に選択する

**Modified method 2:** 類似度として2次構造を考慮しない(配列のみの)類似度を使用する(実験ではclustalwのスコアを使用)

**Modified method 3:** 「擬クリーク」ではなく「クリーク」を抽出する

	Proposed		Method 1		Method 2		Method 3	
	Relevance	Recovery	Relevance	Recovery	Relevance	Recovery	Relevance	Recovery
Dataset 1	<b>0.711</b>	<b>0.846</b>	0.201	0.301	<b>0.711</b>	0.687	0.331	0.5
Dataset 2	0.716	<b>0.837</b>	0.307	0.362	<b>0.722</b>	0.822	0.35	0.525

- オーバーラップする2次構造を全て使用したほうが**断然性能がよい**
- 類似度として2次構造を考慮したほうが**若干性能が良い**
- クリークを抽出するよりも擬クリークを抽出したほうが**断然性能が良い**

(4) 成果の活用と公開

手法から予想されるとおり、RNAmineは並列化を行うことが難しかったが、今回提案する手法は並列化が容易な手法となっている。RNAcliqueを用いた解析例については後述する。本開発は、IP SJ Transactions on Bioinformatics Vol.2, pp.36-46, 2009で論文発表した。

### 2. 1. 1. 1. 3. CentroidFold (RNA 配列からの 2 次構造予測ソフトウェア)

#### (1) 背景

単一の RNA 配列からその 2 次構造を正確に予測することは、バイオインフォマティクスにおける古典的な問題の一つである。その重要性は、近年の機能性 RNA の発見により増してきている。

#### (2) 手法

2 次構造の確率分布（本研究においては与えられているものとする）に対して、 $\gamma$  セントロイド推定量と呼ばれる推定寮を用いて 2 次構造予測（デコーディング）を行う手法を開発した(図 4)。  $\gamma$  セントロイド推定量は、 $\gamma$  TP+TN の期待値を最大にするような推定量（期待精度最大化推定量）である。RNA の 2 次構造中の塩基対を正確に予測することは、生物学的に重要であるため、2 次構造予測の評価には、塩基対に関する SEN、PPV、MCC が通常利用されているが、 $\gamma$  セントロイド推定量は、ML 推定量（Mfold や RNAfold で使用）および MEA 推定量（CONTRAFold で使用）に比べ、これらの評価指標に適した推定量であることを理論的に示すことが可能である。CentroidFold は一つのパラメタ  $\gamma$  を有するが、この  $\gamma$  を調整することにより、予測 2 次構造の SEN と PPV を調節することが可能である。

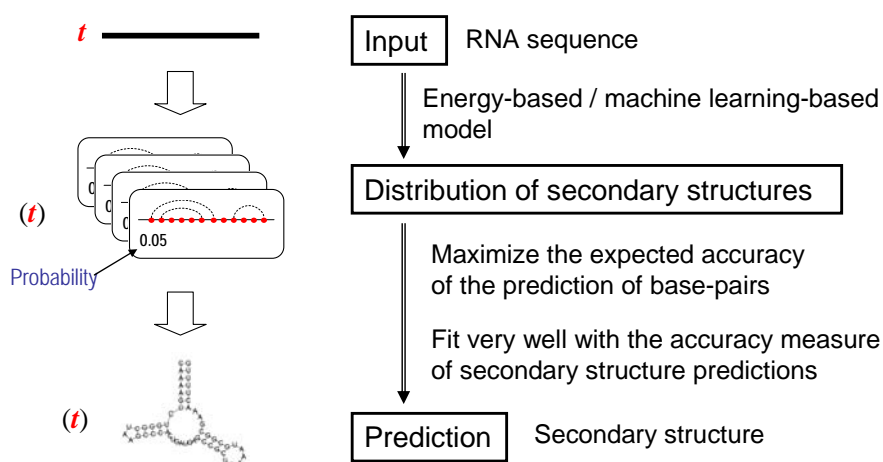


図 4. CentroidFold アルゴリズムの概要

#### (3) 計算機実験

標準的なベンチマークデータセットである S151-Rfam データセットを用いて、評価実験を行った。CentroidFold は既存の 2 次構造予測ツールの中で最も精度が良いことが確かめられた(図 5)。

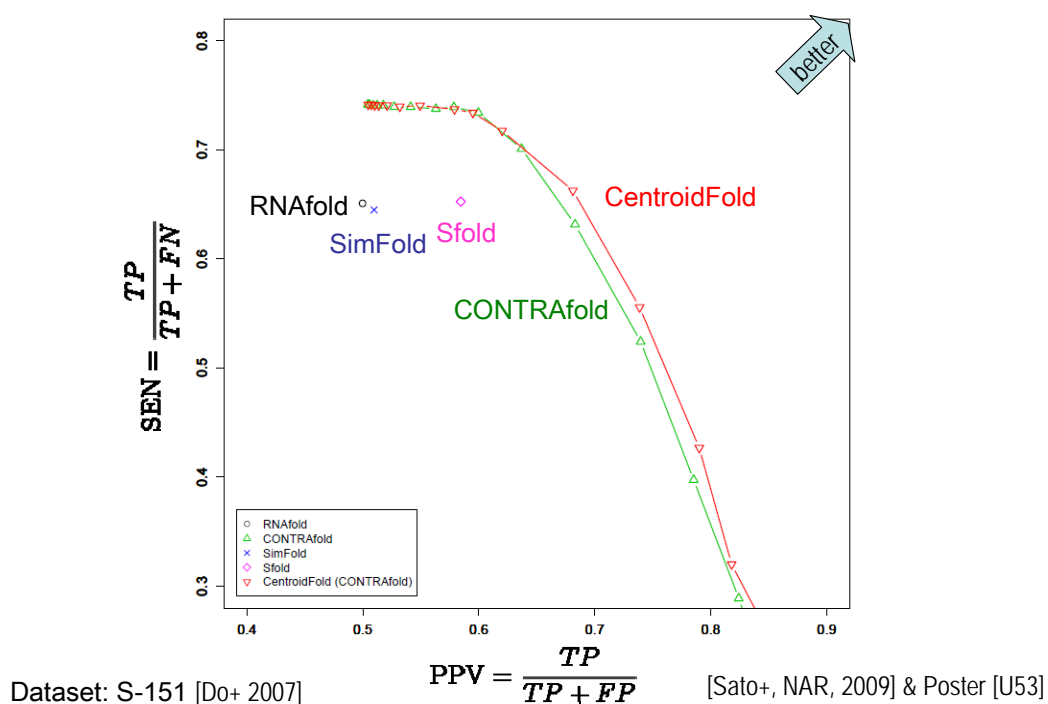


図5. 2次構造予測の性能

#### (4) ディスカッション

MCCを最大化する推定量は、効率的に計算することが不可能である。それどころか、与えられた2次構造に対して、MCCの期待値を計算することも一般には難しい(全ての2次構造候補を数え上げる必要がある)。我々は、偽期待MCCと呼ばれる、効率よく計算可能な量を提案し、偽期待MCCが期待MCCの非常に良い近似となっていることを発見した。

この、偽期待MCCと $\gamma$ セントロイド推定量を組み合わせることにより、小さな計算のオーバーヘッドにより、SENとPPVのバランスの取れた2次構造を予測可能な手法を提案している。

#### (5) 成果の活用と公開

CentroidFoldは現在最高精度の2次構造予測ツールである。WEBサーバ(図6)およびソフトウェアは無償で利用可能である。ソフトウェアについてはHamada et al., *Bioinformatics* 25(4), pp.465-473, 2009で、ウェブサーバーについてはSato et al., *Nucleic Acids Res.* 37(Web Server), pp.W277-W280, 2009でそれぞれ論文発表した。

<http://www.ncrna.org/centroidfold>

The screenshot shows the CentroidFold web interface. On the left is a navigation menu with categories like 'Home', 'About', 'FAQ', 'Contact', 'Links', 'Tools', 'Publications', 'Software', 'Database', 'Help', and 'Privacy Policy'. The main content area has a header 'CentroidFold' and a brief description: 'CentroidFold predicts an RNA secondary structure from an RNA sequence. FASTA and one-sequence-in-a-line format are accepted for predicting a secondary structure per sequence. It also predicts a consensus secondary structure when a multiple alignment (CLUSTALW format) is given. Currently, the input sequence should be less than or equal to 500 bases. For more information, have a look at the help page or click directly on the input field.' Below this is a large empty text box for input, labeled '[入力]'. Underneath the input field are 'advanced options' and a 'Submit' button. The output area, labeled '[出力]', shows a predicted RNA secondary structure as a 2D diagram with nucleotides represented by colored circles (A, U, G, C) and lines indicating base pairs. A color scale at the bottom of the output window ranges from 0.0 to 1.0.

図6. CentroidFoldのWEBインターフェース

## 2. 1. 1. 1. 4. CentroidHomfold (相同配列群の情報を利用した RNA 配列からの 2 次構造予測ソフトウェア)

### (1) 背景

CentroidFold を開発したことにより、我々は現在最高精度の 2 次構造予測ツールを得ることができた。2 次構造予測の精度をさらに向上させるためにはどのようにすればよいだろうか？一般に、2 次構造予測を行いたいターゲット RNA 配列と同時に、その配列に相同な配列群が同時に利用可能な場合が多い。本研究では、ターゲット配列の 2 次構造予測の精度をさらに向上させるために、相同配列群の情報を効果的に利用した 2 次構造予測手法を開発した(図 7)。

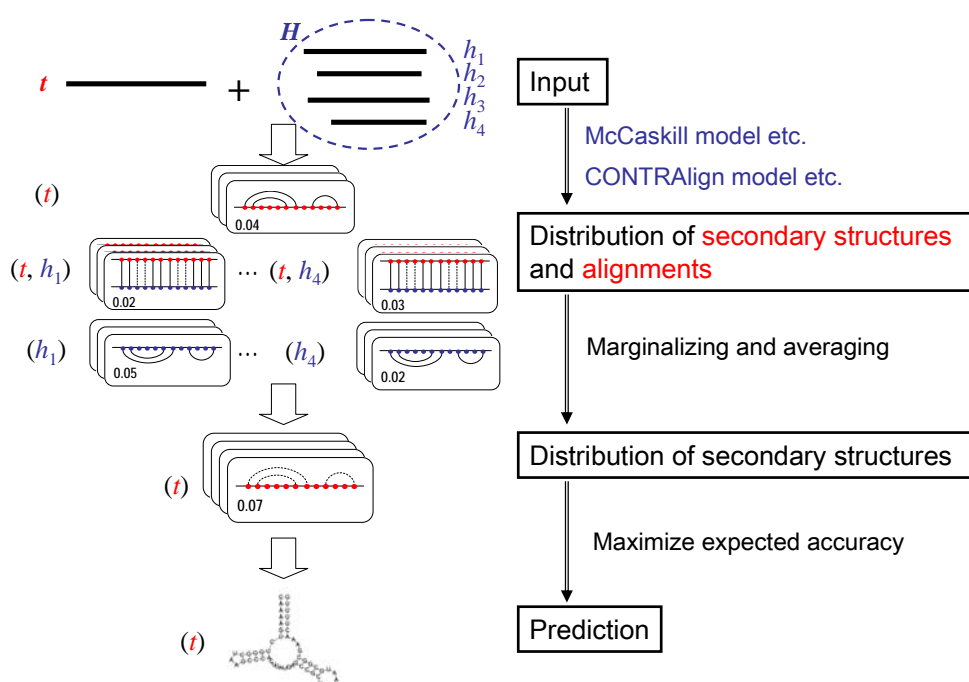


図 7. CentroidHomfold のアルゴリズムの概要

### (2) 手法

ターゲット配列と相同配列の間の可能な構造アラインメントの分布をターゲット配列の 2 次構造に周辺化した分布を考える。ターゲット配列と相同配列が何らかの共通 2 次構造を有すると仮定すると、この分布は、通常のターゲット配列の 2 次構造分布に比べて、優れた分布を与えると考えられる。したがって、この分布に対して  $\gamma$  セントロイド推定量を用いて予測される 2 次構造は、CentroidFold の予測する 2 次構造よりも、正確に 2 次構造を予測可能であることが期待される。しかしながら、この予測方法は多大な計算量を必要とする。従って本研究では、ある種の近似を用いて、高速に 2 次構造を予測する手法を提案し、ソフトウェア CentroidHomfold として実装を行った。CentroidHomfold は、ターゲット配列および各々の相同配列群の可能な 2 次構造全体の分布、および、ターゲット配列と各相同配列との間の可能なアラインメント全体の分布を考慮した手法となっている。

## (3) 結果

CentroidHomfold は相同配列群の情報を効果的に利用することにより、従来の2次構造予測の精度を大幅に向上させることがわかった(図8)。

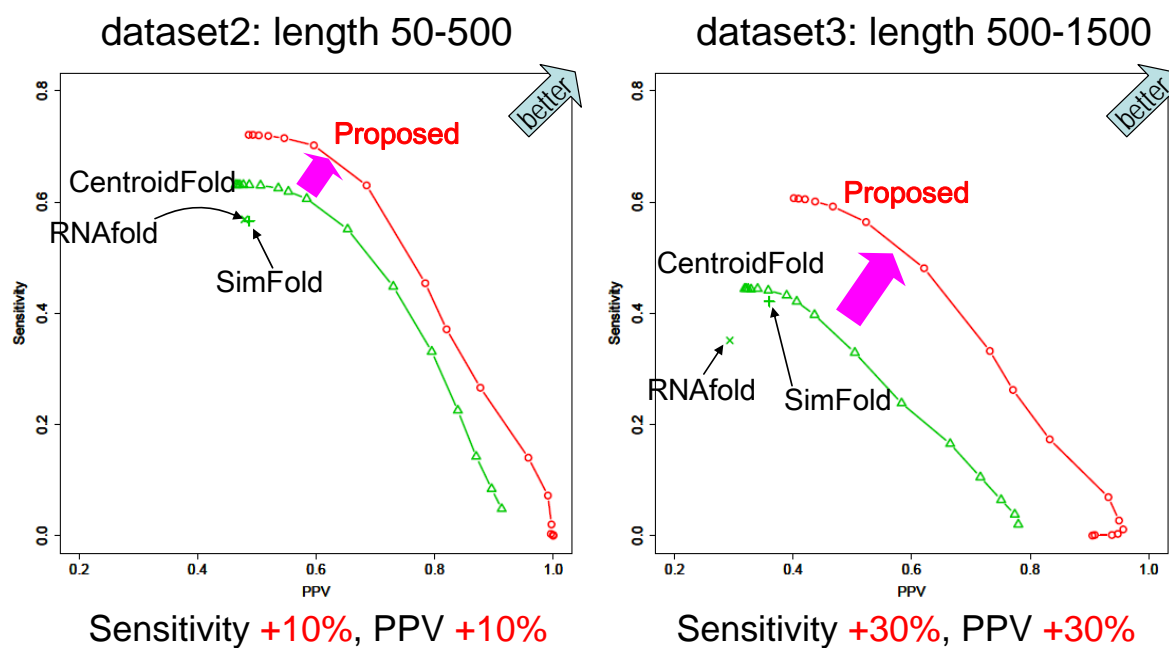


図8. 相同な配列群の情報を利用することによる2次構造予測の精度の大幅な向上

## (4) 成果の活用と公開

CentroidHomfold については、ISMB2009 で口頭発表し、Hamada et al., Bioinformatics 25(12), i330-i338, 2009 に論文が採録された。

## 2. 1. 1. 1. 5. CentroidAlifold (RNA マルチプルアラインメントからの共通 2 次構造予測ソフトウェア)

### (1) 背景

与えられた RNA のマルチプルアラインメントから、そのアラインメントに共通する 2 次構造を予測することは、比較ゲノムによる機能性 RNA の探索や RNA の系統解析において広く用いられる基本的な手法である。

### (2) 手法

アラインメントに含まれる各 RNA 配列の 2 次構造を出来る限り正確に予測する推定量を設計し、CentroidAlifold として実装を行った。CentroidAlifold で利用されている推定量は RNAalifold で利用されている推定量や McCaskill-MEA, PETfold で利用されている推定量に比べ理論的に優れていることを示すことが可能である。

### (3) 計算機実験

他の手法に比べて、共通 2 次構造予測の精度が優れていることが示された (図 9)。

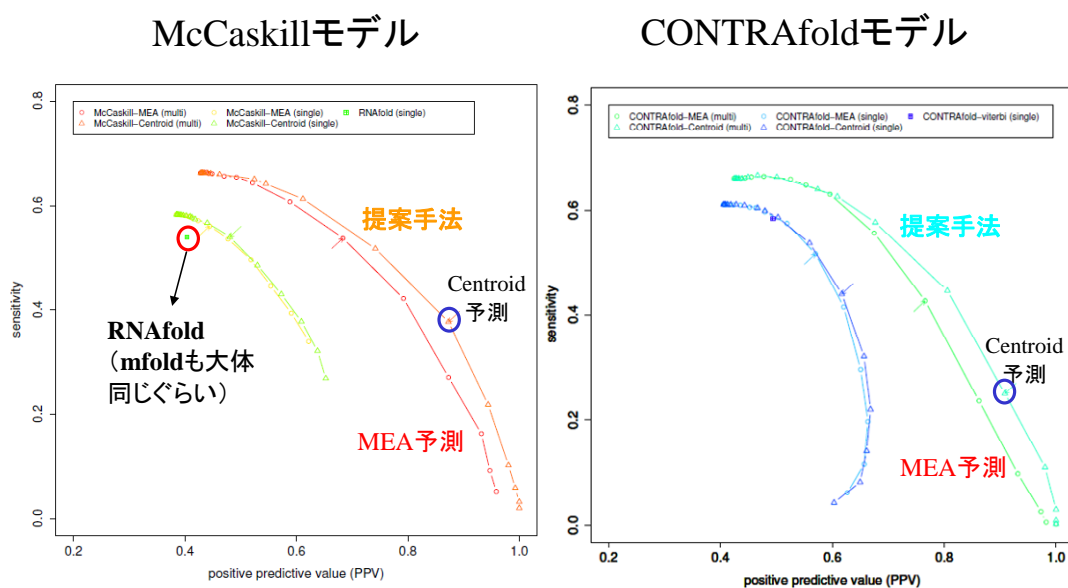


図 9. CentroidAlifold の性能評価



## 2. 1. 1. 1. 6. CentroidAlign (2次構造を考慮した RNA マルチプルアラインメントソフトウェア)

### (1) 背景

2次構造がその機能に密接に関連している機能性 RNA の配列群をマルチプルアラインメントする場合、通常のアラインメントツールのように、配列情報のみを考慮するだけでは不十分であり、各 RNA 配列の2次構造も同時に考慮したアラインメントを行う必要がある。しかしながら、2次構造を考慮したアラインメントは、一般に、膨大な計算量を必要とし、現実的な手法ではない。

### (2) 手法

期待精度最大化原理に基づき、2次構造情報を考慮して、アラインメントのカラムを正確に予測可能な推定量を提案した(図 10)。この推定量は、構造アラインメント上の分布を通常のアラインメント空間に周辺化した分布に対する  $\gamma$  セントロイド推定量に基づいている。実際にこの推定量を計算するためには多大な計算量を必要とするため、実際には、近似を行い高速でアラインメントを計算できるように工夫を行っている。

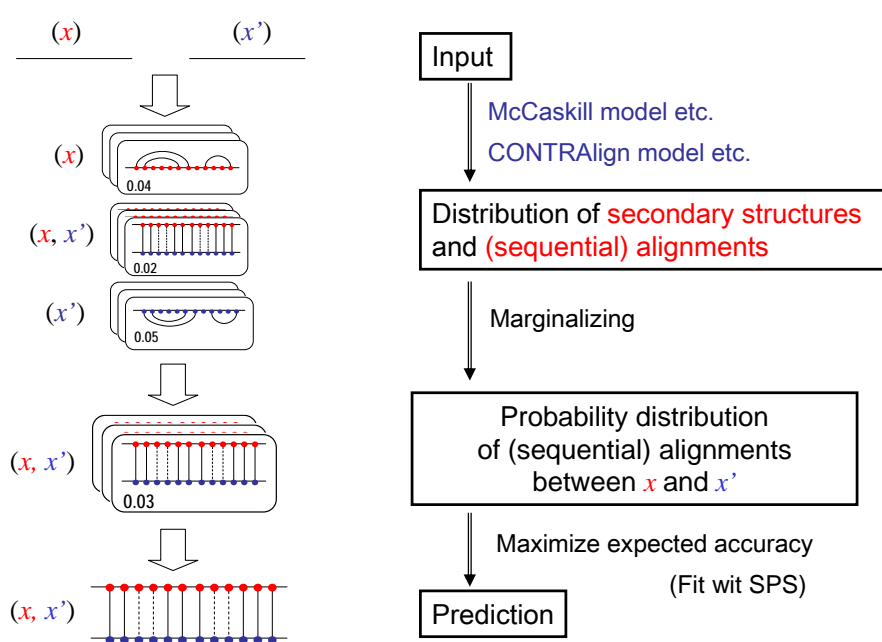


図 10. CentroidAlign アルゴリズムの概要

## (3) 計算機実験

5つのベンチマークデータセットを用いた実験では、CentroidAlignfoldは他のツールに比べて高速かつ正確にアラインメントを予測することが可能であることがわかった(表2)

表2. CentroidAlignfoldの性能評価

Table 3. MXSCARNA dataset

aligner	n	SPS	SCI	SEN	PPV	MCC	TIME
contralign	1693	0.79	0.58	0.64	0.67	0.63	707
probcons	1693	0.78	0.54	0.63	0.66	0.63	487
centroid_align	1693	<b>0.80</b>	0.67	0.70	0.69	0.68	<b>2000</b>
lara	1693	0.77	0.71	0.70	0.68	0.68	61694
mafft-xinsi	1693	<b>0.80</b>	0.71	0.72	0.69	0.69	4316
mlocarna	1693	0.77	<b>0.80</b>	<b>0.75</b>	0.68	0.70	468792
murlet	1693	0.79	0.63	0.71	0.69	0.69	500469
mxscarna	1693	0.78	0.69	0.73	<b>0.70</b>	0.70	2540
raf	1693	0.79	0.72	<b>0.75</b>	<b>0.70</b>	<b>0.71</b>	41078
rcoffee	1693	0.78	0.61	0.67	0.68	0.66	4822
stral	1693	0.74	0.57	0.61	0.63	0.60	2021

See the caption in Table 1

Table 4. MASTR dataset

aligner	n	SPS	SCI	SEN	PPV	MCC	TIME
contralign	52	0.87	0.72	0.64	0.77	0.69	31
probcons	52	0.87	0.72	0.64	<b>0.78</b>	0.69	<b>18</b>
centroid_align	52	<b>0.88</b>	0.75	0.65	0.77	0.70	47
lara	52	0.86	0.77	0.69	<b>0.80</b>	<b>0.73</b>	3620
mafft-xinsi	52	<b>0.88</b>	0.78	0.68	0.78	0.71	133
mlocarna	52	0.85	<b>0.80</b>	0.68	0.75	0.71	2117
murlet	52	0.87	0.74	0.67	0.78	0.71	4149
mxscarna	52	0.86	0.73	0.66	0.77	0.70	50
raf	52	0.86	0.74	<b>0.70</b>	0.77	<b>0.73</b>	272
rcoffee	52	0.87	0.74	0.66	0.78	0.70	223
stemloc-ama	51	0.86	0.72	0.63	0.77	0.68	353453
stral	52	0.81	0.70	0.61	0.75	0.65	<b>32</b>

See the caption in Table 1

CentroidAlign

- one of the fastest tools (that consider secondary structures)
- better SPS (sum-of-pairs score) than other tools
- better SCI,SEN,PPV,MCC than CONTRAlign and probcons

## (4) 成果の公開と活用

CentroidAlignについては、Hamada et al., Bioinformatics 25(24), pp. 3236-3243, 2009に論文発表した。

## 2. 1. 1. 1. 7. RNAmine を用いた機能性 RNA 候補領域の発見

RNA 配列群からの共通 2 次構造パターン抽出プログラム RNAmine を用いて、ヒトゲノムの intergenic 領域とイントロンから、配列保存度が高い 100 塩基以上の配列を抽出し解析を行った。Intergenic 領域に対しては、平成 18 年度の解析で EST のサポートを課していた処理を省いた 32,474 配列のデータセットを解析対象とした。イントロンに対しては、平成 18 年度の解析で得られた 10,609 配列のデータセットを解析対象とした。これらの大規模データを直接 RNAmine の入力データとすることは現状では困難なため、クラスタリングによる前処理を行うことでクラスター単位での解析を行った。まず、BLASTClust で作成した全てのクラスターに対して RNAmine でモチーフ抽出を行った。RNAmine の出力結果を RNAforester でアライメントし、RNAz に入力したときの計算結果の内、特にスコアが高かったものを RNA ファミリー候補とした (図 11)。解析の結果、構成メンバーがユニークな配列からなる 25 モチーフを抽出することに成功し、

133 箇所の新規機能性 RNA 候補を得た。これらの機能性 RNA 候補に対してマイクロアレイによる実験を進めた。

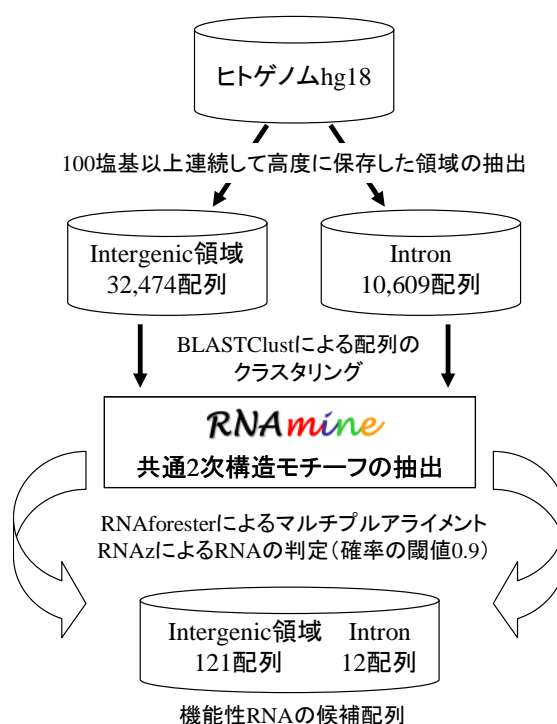


図 1 1. RNAmine を利用した機能性 RNA 候補配列の探索

RNAmine で抽出した配列から設計したプローブのマイクロアレイ実験結果に対して、Hela、Brain、Testis、Thymus のいずれかの組織で 50 以上の信号強度が記録され、組織特異的な発現が見られたプローブについて解析した。解析の結果、RNAmine で作成した同一クラスターから設計されたプローブの内、Brain の信号強度比が他の 3 組織の 2 倍以上であったプロ

ープを2プローブ得ることが出来た（表3）。

**表3. RNAmineの同一クラスターから設計されたプローブ**  
**（プローブ番号6476と6479がbrainで信号強度比2倍以上）**

プローブ 番号	信号強度比				測定値			
	Hela	Brain	Testis	Thymus	Hela	Brain	Testis	Thymus
6475	1	1.34	2.02	2.02	29.743	25.8569	69.19543	37.40676
6476	1	12.61	0.89	1.08	17.51814	142.8838	17.96623	11.78372
6479	1	2.31	0.6	0.42	109.0125	162.6364	75.73906	28.57087
6482	1	0.47	1.19	0.6	71.18672	21.75913	97.28877	26.64562

### 2. 1. 1. 1. 8. RNAclique を用いた機能性 RNA 候補の発見

実データを用いた大規模な解析を2例行った。まず、第1の解析では、Human ゲノム中で43個の既知 miRNA を含む遺伝子間領域 (95,757 塩基) に対して手法を適用した。その結果、スコアが1位のクラスタ (66 配列を含む) に既知の43個の miRNA と1個の Berzikov による予測 miRNA が含まれていた。クラスタに含まれる残りの配列は新規 miRNA の可能性もある。第2の解析では、Washietl らにより予測された機能性 RNA 候補領域 (71,970 領域) に対して手法を適用した。その結果、492 のクラスタが抽出された。入力データ中で5本以上存在する既知の機能性 RNA ファミリー SECIS, SNORA70, let-7, mir-10, mir-17, tRNA の各々に対して、カバレッジが最も高かったクラスタのカバレッジの値は、SECIS (0.17)、SNORA70 (0.92)、let-7 (1)、mir-10 (0.80)、mir-17 (0.80)、tRNA (0.53) であった。Washietl らのデータセットを用いた場合の計算時間は22CPUで64,708秒であった。このように新しく提案した手法は大規模な配列群に対して適用することが可能な手法である。

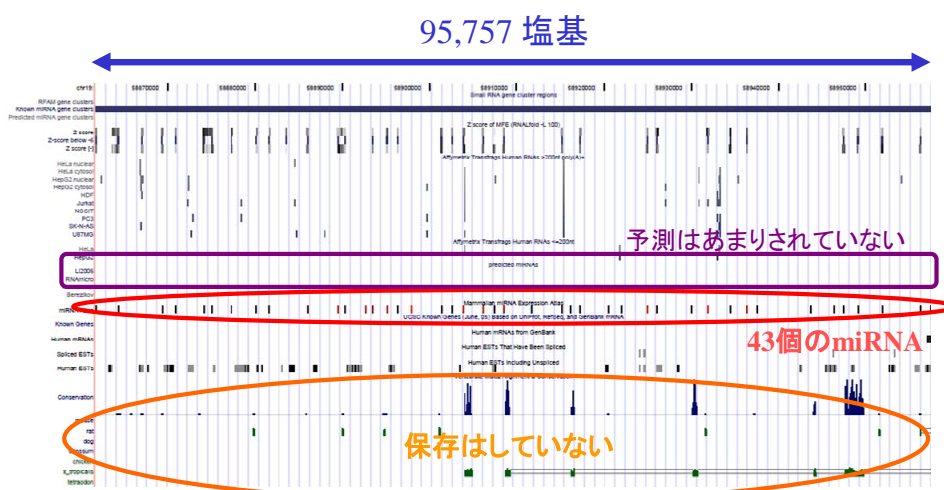


図12. 43個の miRNA が存在する Human ゲノム中の領域

一方、上記の解析例でも明らかになった通り、2次構造の類似性のみを指標としてクラスタ抽出をおこなった場合、多数のクラスタが発見されることがわかった。そのため、さらに確度の高いクラスタを抽出するため、(a) タイリングアレイ発現データ [Kapranov+2007] とのオーバーラップ、(b) EST とのオーバーラップ、(c) Transposon Free Region [Simons+2005] とのオーバーラップ、(d) Most Conserved Region [Siepel+2005] とのオーバーラップ、(e) Indel Conserved Region [Lunter+2006] とのオーバーラップ、(f) Low Z-score Region とのオーバーラップ、RNAz により予測された機能性 RNA 候補 [Washietl+2006] とのオーバーラップ、(g) Evofold により予測された機能性 RNA 候補 [Pedersen+2006] とのオーバーラップの9個の特徴を同時に満たすクラスタの効率的かつ完全な探索を行う手法の実装も行った。(a)を満たし、(b)-(g)の中の3つ以上を共通に満たす、2次構造が類似した

クラスタを探索した結果、(a), (b), (c), (d)の特徴を共通に有するクラスタ (10 配列)が発見された (図 1 3)。このクラスタは2次構造が類似しているだけでなく、その他の特徴も共有しているためより確度の高いクラスタであると考えられる。



図 1 3. RNAcliqueにより得られたクラスタ

## 2. 1. 1. 1. 9. まとめ

本研究開発項目においては、機能性 RNA のための総合的なバイオインフォマティクス技術の開発を行った。その結果、RNA の2次構造予測ツール CentroidFold や2次構造を考慮した高速なマルチプルアラインメントツール CentrodAlign に代表される多数の基本的なソフトウェアを開発することができた。特に、CentroidFold は、現時点において、最高精度の2次構造予測精度を有している。これらの基本的なツール群は、今後、機能性 RNA を解析する際のツールとして広く利用されることが期待される。

## 2. 1. 1. 2. 高速な RNA 配列の比較検索技術の開発

### (1) RNA配列の高速な構造アラインメントと共通2次構造予測

機能性 RNA においては配列の類似性だけでなく、その二次構造をみなければ分類や機能推定が難しいが、RNA 配列からの二次構造予測の制度には限界があり、二次構造を考慮した配列の比較・整列技術も実用的な既存技術も従来は存在しなかった。

RNA 配列の潜在的な共通二次構造と配列類似性を同時に考慮しながら厳密に整列させるためには、長さ  $L$  の配列  $N$  本に対して、 $O(L^{3N})$  の計算時間と  $O(L^{2N})$  の記憶容量が必要である (Sankoff アルゴリズム) ことが知られている。これでは、2 本の配列の整列であっても、10 倍長い配列に対しては 100 万倍の計算時間がかかり、数百塩基程度の配列の整列に必要なメモリが通常の計算機のメモリ量をはるかに超えてしまうため、ゲノムレベルの解析に利用することは全く不可能である。そのため、現実的な計算時間と記憶容量で配列の比較・整列を行うため、主に次の 3 種類のアプローチが有望である。

- (a) Sankoff アルゴリズムの動的計画法において、整列させる範囲の限定、構造の分岐の粒度の制限、塩基対間の距離の制限を組み合わせることで高速化する方法。
- (b) 二次構造を形成するステム領域の 5' 側と 3' 側の整列を独立に行い、ステムの整合性は後処理することによって大幅に計算時間と記憶容量を節約する方法。
- (c) 各配列から抽出された二次構造を考慮した特徴量の比較を、カーネル法を用いて効率的に計算することにより、配列を直接整列させることを回避する方法。

本研究課題では、主に (a)、(b) のアプローチを用いた。(c) のアプローチは、次の研究課題 2-3 確率文法とカーネル法による RNA 配列情報解析技術の開発で用いられた。

(a) のアプローチによって、200 塩基程度の配列群に対する Sankoff アルゴリズムが通常の計算機で実行できるソフトウェア Murlet (<http://murlet.ncrna.org>) を開発し、整列の精度で世界最高性能を達成した (Kiryu et al., Bioinformatics 23(13), pp. 1588-1598, 2007)。動的計画法の領域制限を領域の幅によって行うと、その幅を配列長の差以上に取らなければならないため、配列群の配列長が均一でないと著しく効率が低下する。そこで、MEA 原理に基づくアラインメントを基準として動的計画法の領域を設定して効率化した。同時に二次構造モジュールの分岐箇所の粒度を制限することによっても大幅に高速化した。

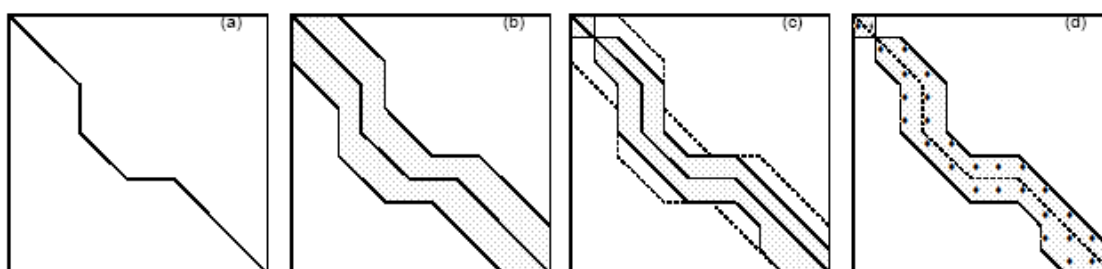


図1. (DP 領域の制限。(a) MEA 原理に基づく初期アラインメント (b) 初期アラインメント近傍に制限された DP 領域 (c) 一致確率の低い部分を削った DP 領域 (d) 制限された分岐可能場所を点で示した

(b)のアプローチによって、数千塩基の RNA 配列を二次構造を考慮して整列させる高精度なソフトウェア SCARNA を開発した。2本の配列の比較・整列では現実的な計算時間で実行できない Sankoff アルゴリズムを除けば世界最高性能を達成した (Tabei et al., *Bioinformatics* 22(14), pp.1723-1729, 2006)。さらにアルゴリズムを配列群の多重整列に拡張した MXSCARNA (<http://mxscarna.ncrna.org>) を開発し、世界最高速と最高性能を同時に達成した (Tabei et al., *BMC Bioinformatics* 9:33, 2008)。さらに、RNA 配列群から二次構造と配列が局所的に類似した配列を抽出してアラインメントを行う Scarna-lm を開発した (Tabei et al., *Bioinformatics* 25(12), pp.1498-1505, 2009)。これらの手法では、MaCaskill アルゴリズムによって得られた各 RNA 配列の塩基対確率行列からステム領域の候補を抽出し、その固定長の構成要素の 5' 側と 3' 側を技巧的な動的計画法によって整列させる。多重整列においては配列群の塩基対確率行列を平均化することによってロバストな共通ステム候補を抽出できることが明らかとなった。



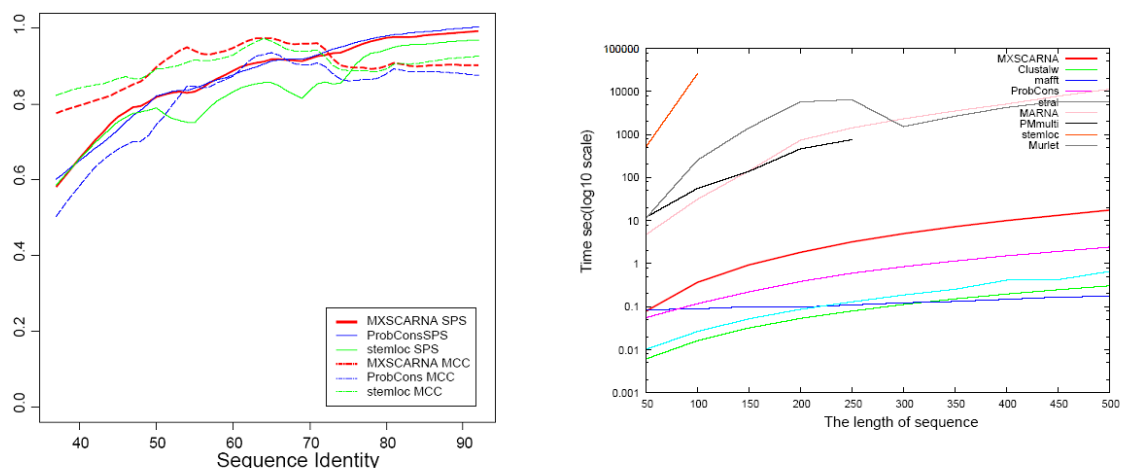


図 2. MXSCARNA による多重構造アラインメントの精度と計算速度。(左図) SPS (Sum of Pairs Score) および MCC (Mathew' s Correlation Coefficient) による多重アラインメントの精度と配列相同性の関係。MXSCARNA が両方の指標で安定した精度を示している。(右図) 配列長と計算時間の関係。二次構造を考慮したアラインメント手法では MXSCARNA が唯一、大規模解析に適用可能であることがわかる。

ゲノム中からの機能性 RNA 発見のためには、ゲノムブラウザ等から得られる RNA のマルチプルアラインメントから共通 2 次構造を求める手法が必要である。本開発では、バイオインフォマティクスにおいて重要性を増している期待精度最大化 (MEA: Maximum Expected Accuracy) 原理に基づく手法 (McCaskill-MEA) を開発し、従来よりも正確な共通 2 次構造予測を実現した (Kiryu et al., Bioinformatics 23(4), pp.434-441, 2007)。

## (2) 高度な動的計画法による 2 次構造解析手法の開発

RNA 分子は、配列上離れた位置の塩基同士が 2 次構造上で塩基対を形成する可能性があるため、RNA 分子の始点と終点を定めなければ正確な 2 次構造予測を行うことができない。しかし、RNA 分子として転写された場合の潜在的な 2 次構造を考慮した網羅的な解析をゲノム配列に対して行う場合には、あらかじめ RNA 分子の範囲を決めることは困難である。ゲノム上に点在する潜在的な機能性 RNA 候補や、miRNA のターゲットとなる mRNA など様々な配列に対して、2 次構造を考慮した網羅的な解析を行うためには、始点と終点を特定せずに局所的な 2 次構造を解析する手法が必要である。

一方、RNA の 2 次構造予測に用いられてきたアルゴリズムは配列長の 3 乗に比例した計算時間と 2 乗に比例した記憶容量を必要とするから、長いゲノム配列に対して直接このアルゴリズムを適用することは明らかに不可能である。従来は、決められた固定長の窓を少しずつずらしながら計算された塩基対確率を平均化することによって近似的な塩基対確率が

計算されていた。

本開発では、塩基対を形成する最大の塩基間距離  $W$  (固定) と、配列長  $L$  (一般に長い) に対して、高度な動的計画法を駆使することによって  $LW^2$  に比例した計算時間と  $L+W^2$  に比例した記憶容量で正確な塩基対確率が計算できることを示し、Rfold というソフトウェアとして実装、公開した。また、Rfold で計算された塩基対確率に基づく局所 2 次構造予測は、従来の手法よりも正確であることを示した (Kiryu et al., Bioinformatics 24(3), pp. 367-373, 2008)。

さらに、Rfold で用いたアルゴリズムを発展させて、ゲノム内に潜在的に存在する RNA の二次構造的な特徴を網羅的に計算するアルゴリズムを Raccess プログラムとして実装し、これを用いてヒトゲノムの統計解析を行った。

siRNA や miRNA などの機能性 RNA は、mRNA と相補的に結合することで、その機能を発揮する (図 3 上)。このとき、もしターゲットとなる領域が強い二次構造をとるならば、それが障害となり機能を発揮できない (図 3 下) ため、miRNA や siRNA による抑制効果が重要な遺伝子においては、ターゲット領域が強い二次構造をとらないように、進化的な圧力が働いている。最近になって、ターゲット領域が二次構造を組みにくいアクセシブルな領域であるという性質を考慮することが miRNA の予測や、siRNA の設計に有効であることが知られるようになってきた。

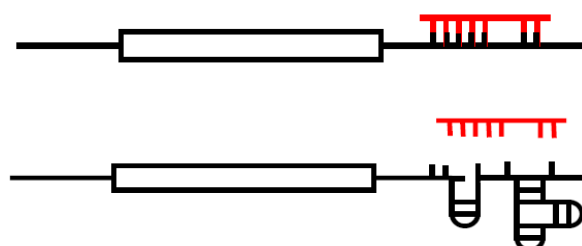


図 3. mRNA(黒)、miRNA(赤)

これまでの二次構造研究は、塩基対を形成する部分に注目するものが多く、塩基対を組まない領域の性質についての研究は非常に少なかった。しかし、siRNA や miRNA の例から類推すると、転写物の中でアクセシビリティが特に強く保存された領域は、なんらかの機能性 RNA のターゲットになっている可能性があると考えられる。そこで我々は、ゲノムスケールでアクセシブルな領域を計算することのできるアルゴリズムを開発し、Raccess というソフトウェアとして実装した。

このプログラムをヒトゲノム全体に対して適用し、潜在的に二次構造を組まない進化圧力が働いている領域を網羅的に探索 (図 4) し、その統計解析を行った。その結果、ゲノム上の転写領域には全般的に二次構造に対する拘束が非常に強く働いていることが明らかになった。しかも、転写物の機能の異なる各部位において、進化的拘束のパターンも様々

であることが分かった。

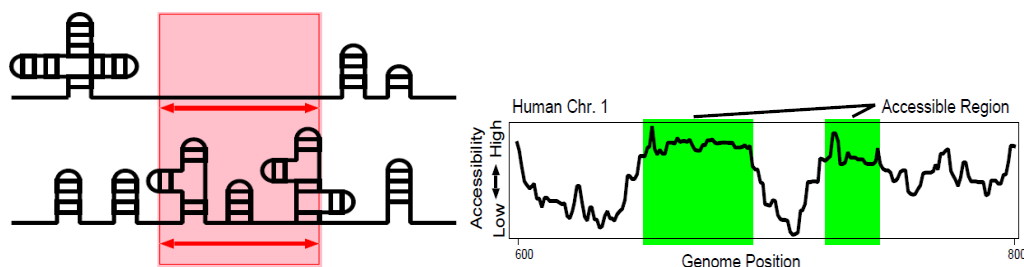


図4. ゲノム上の位置とアクセサビリティ

(左) Raccess はRNAの各範囲について、そこがアクセサブルである確率を計算する。

(右) ゲノムの各位置について計算されたアクセサビリティ。

緑の領域は他のゲノム領域に比べて特にアクセサブルであることを示す。

例えば、タンパク質のコード領域は、全領域についてアクセシビリティを保つ拘束が働いているが、これは、mRNAの翻訳がスムーズに進むためであると予想される。5' UTR領域には、二次構造を作ろうとする進化的圧力が働いているが、これはこの領域が翻訳制御に関わる二次構造をとることが多いこととコンシステントである。3' UTR領域は、全般的にはアクセシビリティが低いですが、ストップコドンの下流150塩基位までは、アクセシビリティが高い。これは、翻訳がスムーズに進むためであるとも考えられるが、miRNAのターゲットサイトが、ストップコドンの近くにあるほど抑制効果が高いこととも関連があると考えられる。このほかに、mRNA以外の転写領域として、LINEやSINEなどのリピート配列についても統計解析も行った。すると、トランスポゼースをコードしているL1などはmRNAと似たアクセシビリティの特徴を持つことが分かった。一方ヒトゲノム最大のリピートファミリーであるAlu配列は、強固な二次構造を保存するような圧力が働いており、細胞の中で、構造RNAとして生理機能をもつものが相当数あることが予測された。

この一連の統計解析で、一番大きな発見は、イントロン領域には二次構造を保存する強い進化的圧力が働いていることを初めて検出したことである(図5)。これまでのゲノム解析では、イントロン部と遺伝子間領域を区別するような配列特徴は知られていなかったが、我々の解析では、pre-mRNAのイントロン部の全領域に渡って、強固な二次構造をとるような非常に強い進化的圧力が存在することが分かった。この特徴は、一義的には長いイントロンが効果的にスプライシングされるために生じたと考えられるが、イントロンが、進化の過程で機能性構造RNAを供給するreservoirとして働いている可能性も考えられる。

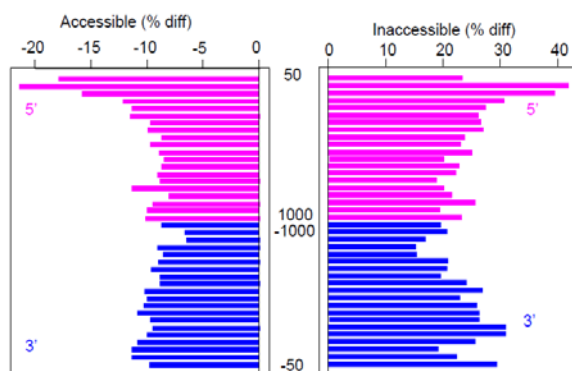


図5. イントロン領域のアクセサビリティ。左図、右図は、イントロン領域において、それぞれアクセサブルな領域、アクセサブルではない領域の割合のバックグラウンドからのずれを示す。

縦軸は5'、3' 端からの塩基数（1000塩基まで）を示す。イントロン領域では、アクセサブルでない領域が多いことを示している。

### （3）RNA 二次構造の揺らぎを考慮した配列解析ツール群

1990年代までのRNAの配列解析は、RNAの遺伝子の二次構造を正確に求める方法論が主だったが、高等動物に多数存在する非コードRNAの研究に伴い、RNA配列解析の目的も多様性を増している。プロジェクトの中で、共通二次構造(McCaskill-MEA)、構造RNAのマルチプルアライメント(Murlet)、局所塩基対確率(Rfold)、アクセシビリティ(Raccess)などを計算するアルゴリズムの開発及び、実装を行ってきたが、さらにTurnerのエネルギーモデルから様々な二次構造的特徴を計算するアルゴリズムの開発及び実装を行った

(Rprofile、Renergy、Rpair、Rchange)。まずRprofileは、ゲノムの各領域で、二次構造の各コンポーネント（内部ループ、バルジ、マルチループ）などが出現する確率を計算するプログラムである。Renergyは、ゲノムの各スライディングウィンドウで、その幅のRNAの二次構造エネルギーの期待値を効率的に求めるプログラムである。Rpairは、ゲノムの塩基ペアが、対を組むときと組まないときで平均エネルギーの差を求めるプログラムである。これにより、各塩基のペアが二次構造安定化にどの程度寄与しているかを計算することができる。Rchangeは、RNA配列に塩基置換が入ると平均エネルギーがどの程度変化するかを計算することができる。これらのプログラムに共通するのは、RNAの二次構造の揺らぎを考慮しているという点で、確固とした構造を取らない転写物の解析に役に立つことが期待される。

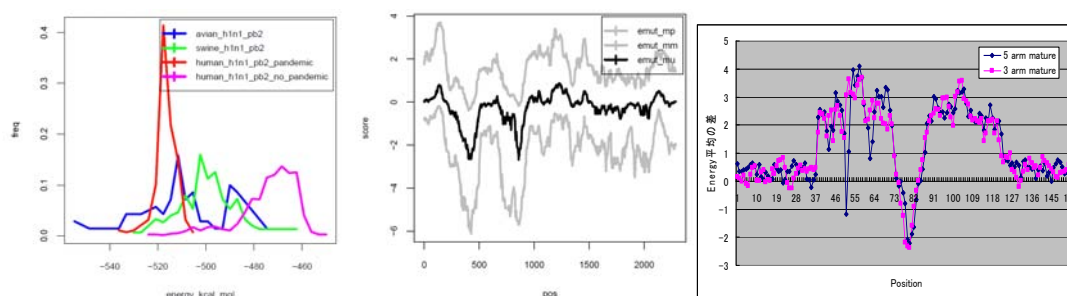


図6. (左)Renergyにより計算した、インフルエンザウィルスのPB2ポリメラーゼ mRNA の二次構造エネルギー分布。2009年のヒトにおけるパンデミック由来のものを除けば、鳥、豚、人と体温が下がるにつれて、mRNAの二次構造エネルギーが弱くなっていくのがわかる。(中) Rchangeによる、PB2ポリメラーゼ mRNA 配列の各位置で塩基置換前後の二次構造エネルギーの変化。400塩基目と800塩基位置のあたりに置換が起こるとFoldエネルギーが弱くなる場所がある。(右) Rpairを用いた pre-miRNA 遺伝子周りのエネルギー変化のプロファイル。5'アームに mature miRNA を持つ pre-miRNA (青) では、mature miRNA の 5' 末端部分が塩基対を組まない傾向がある。

#### (4) まとめ

本研究開発では、高速で正確な構造アラインメントツールである Murlet、MXSCARNA をプロジェクトの前半で開発することに成功し、①-2「ゲノム配列からの機能性 RNA の網羅的予測」に活用することが出来た。また、局所的な塩基対確率と2次構造を求める Rfold、アクセサビリティを計算する Raccessをはじめ、Rprofile、Renergy、Rpair、Rchange など RNA 配列の2次構造に関連した網羅的な解析を可能とするツール群の開発に成功した。これらのツール群は、miRNA、siRNA の標的予測や核酸医薬の開発に必要な基本的な解析ツールとして活用されることが期待される。

## 2. 1. 1. 3. 確率文法とカーネル法によるRNA配列情報解析技術の開発 集中研①、慶應義塾大学（1）

機能性 RNA の大きな特徴は、水素結合で結ばれる塩基対によって形成される二次構造であり、生体内での機能と相関があることが知られている。機能性 RNA の配列解析では、二次構造の自由エネルギー最小化や二次構造をモデル化する確率文脈自由文法などの確率文法に基づく手法が非常に有効であることが知られている。また、様々な分野でカーネル法によるパターン認識が盛んに研究されており、その有効性が明らかになっている。我々のグループでは、これらの手法を組み合わせることにより、画期的な RNA 配列情報解析技術の開発を行った。

### 2. 1. 1. 3. 1. 確率文法による RNA 配列解析

確率文脈自由文法による RNA 二次構造予測では、用いる文法構造が予測精度を大きく左右する。この研究では、ノンパラメトリックベイズ法の一つである階層ディリクレ過程を用いて確率文脈自由文法を拡張することによって、訓練データから最適な文法構造を学習し、その結果、予測精度が向上することを示した (Sato *et al.*, WABI2009 で口頭発表、Journal of Bioinformatics and Computational Biology に採択)。表 1 では、(Dowell *et al.*, 2004) で提案されている 9 種類の RNA 文法 (G1~G8) を用いて、通常確率文脈自由文法 (SCFG) と階層ディリクレ過程で拡張した確率文脈自由文法 (HDP-SCFG) を比較し、すべての場合において提案手法 HDP-SCFG が二次構造予測精度で SCFG を上回ることを示した。とくに、G1, G3, G4, G5 のように SCFG では精度が非常に低い文法でも、HDP-SCFG に拡張することにより大きく精度を向上させることができるという特徴を持つ。

表 1. MCC における確率文脈自由文法 (SCFG) と提案手法 (HDP-SCFG) の比較

grammar	SCFG	HDP-SCFG	grammar	SCFG	HDP-SCFG	grammar	SCFG	HDP-SCFG
G1	0.32	0.61	G4	0.19	0.58	G6s	0.56	0.57
G2	0.59	0.60	G5	0.04	0.58	G7	0.57	0.58
G3	0.42	0.58	G6	0.56	0.57	G8	0.56	0.59

また、本プロジェクト開始前より開発を行っていた RNA ホモロジー探索ツール PHMMTS を拡張し、位置特異的スコア行列を導入したアルゴリズム PSSMTS を開発した。与えられた共通二次構造を元に、整列されていない配列群から EM アルゴリズムを用いて位置特異的スコア行列を学習する。相同性が低い場合においても、PSSMTS は既存の手法よりも高い精度で RNA ホモロジー検索が可能であることを示した (Sato *et al.*, Journal of Mathematical Biology 56(1-2), pp.201-214, 2008)。

## 2. 1. 1. 3. 2. カーネル法による RNA 配列解析

### (1) ステムカーネルの開発

配列群の多重アラインメントを直接計算することなく、形成しうる全ての二次構造を考慮に入れたカーネル関数であるステムカーネルを開発した。カーネル法の代表的な手法であるサポートベクターマシンに適用し、機能性 RNA のファミリー分類において非常に高精度の識別を実現した (Sakakibara *et al.*, *Journal of Bioinformatics and Computational Biology*, 5(5), pp.1103-1122, 2007)。

### (2) ステムカーネルの高速化

ステムカーネルは、サポートベクターマシンによる識別実験では高い識別性能を発揮するが、計算量が膨大であるために実用的ではないという大きな欠点があった。そこでこの欠点を克服するために、塩基対確率行列から非循環有向グラフ (Directed Acyclic Graph; DAG) を構築し、この上でカーネルを計算することによってステムカーネルを高速化する手法を開発した。また、塩基対確率行列をプロファイル化することによって、マルチプルアラインメントに対するステムカーネルを計算する手法を開発した。サポートベクターマシンによる識別実験では、RNAz の学習に用いられている RNA アラインメントのデータセット (12 ファミリー, 7169 アラインメント) を使用し、表 2 に示すように、既存の手法よりも高精度の結果が得られることを確認した。また、カーネル階層クラスタリングでは、Rfam から選んだ配列 (503 ファミリー, 3901 本) を用い、得られたクラスターと Rfam のファミリーとの一致度を ROC スコアで評価した。既存手法の LocARNA では 0.781 に対し、我々の手法では 0.894 となった。これらの結果から、我々が新しく開発した手法は既存の手法よりも構造を持つ機能性 RNA の類似度として信頼性が高いと言える (Sato *et al.*, *BMC Bioinformatics*, 9:318, 2008)。

表 2. サポートベクターマシンによる識別実験の結果

	ROC score	Specificity	Sensitivity
Profile-Profile Stem Kernel (本手法)	<b>1.000</b>	<b>0.997</b>	<b>0.995</b>
Profile-Profile Local Alignment Kernel	0.973	0.995	0.789
RNAz	0.981	0.944	0.927

### (3) BPLA カーネルの開発と線虫ゲノムへの適用

これまで開発してきたステムカーネルよりも高速かつ高精度な BPLA カーネルを開発した (Morita *et al.*, *Nucleic Acids Research* 37(3), pp.999-1009, 2009)。ステムカーネルでは、二本の RNA 配列の塩基対の組み合わせをすべて考える必要があるため、膨大な計算時間が必要であった。BPLA カーネルでは、与えられた RNA 配列の塩基対確率行列を、それ

ぞれの塩基が自身の左側（上流）と塩基対を組む確率、右側（下流）と塩基対を組む確率、どちらも塩基対を組まない確率、という三種類の確率値に要約し、ローカルアラインメントカーネルと同様にしてカーネル値を計算する。既知 snoRNA 配列で学習した本手法を用いて線虫 *C. elegans* ゲノムをスキャンしたところ、図 1 に示すように高い識別性能を確認することができた。さらに、既知 snoRNA には含まれないが確率値の高いものを新規 snoRNA 候補として、候補に対してプライマーを設計し定量的 RT-PCR 実験を行った（図 2）。実験を行った 48 候補中、Ct 値 25 以下（ネガティブコントロールがひとつも発現しない Ct 値）で発現を確認することができたものは 6 候補、Ct 値 26.5 以下（ネガティブコントロールがひとつのみ発現する Ct 値）で確認することができたものは 14 候補であった。これらの Ct 値で得られた候補のゲノム配列に対する p-value はそれぞれ  $3.6e-12$ 、 $2.8e-32$  であり、効率がよい予測とすることができる。

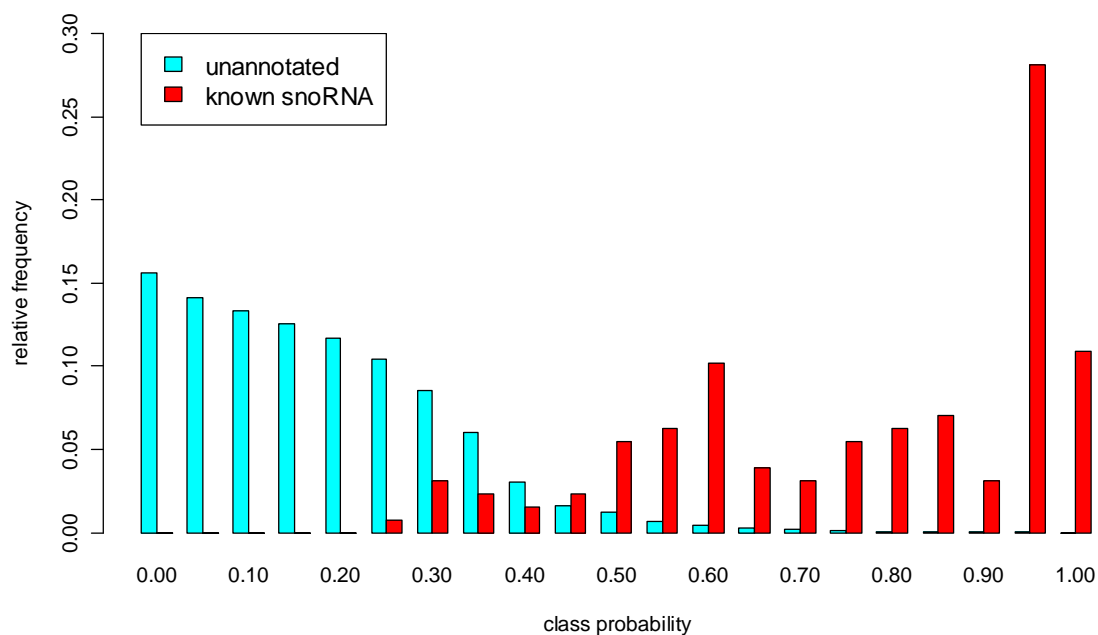


図 1. 線虫 snoRNA のゲノム中における確率値の分布



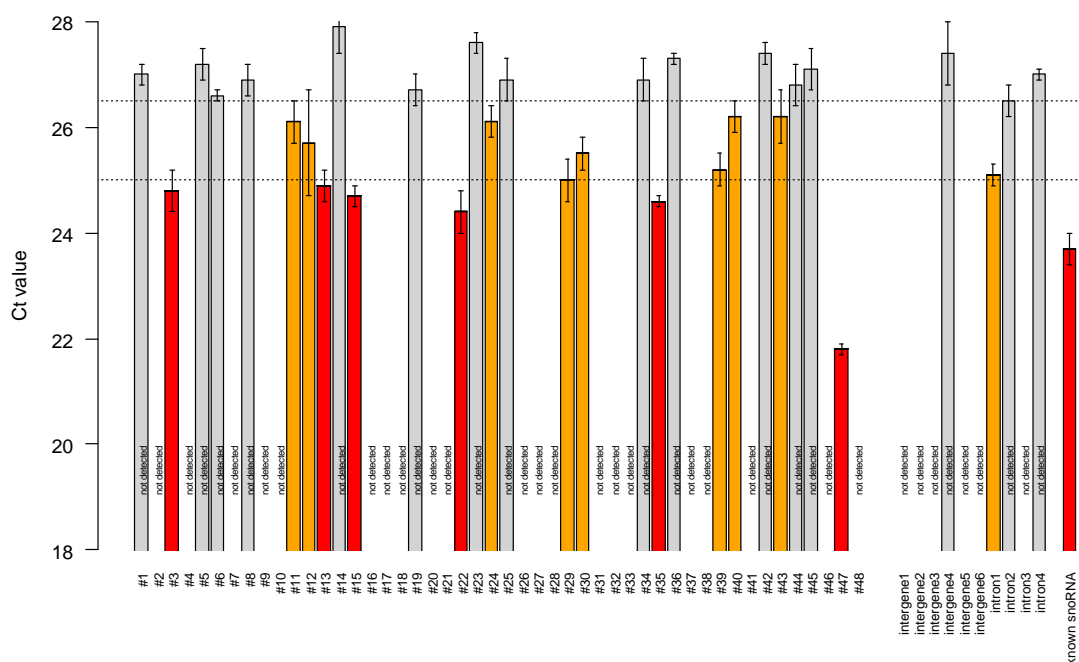


図2. 新規 snoRNA 候補の qRT-PCR による発現確認実験

#### (4) BPLA カーネルのホヤゲノムへの適用

BPLA カーネルによる機能性 RNA 探索と予測候補の発現・機能解析までを一貫して行う実験系の確立を目指し、カタユウレイボヤ *C. intestinalis* ゲノムから snoRNA を探索し、発現確認を行った。まず、C/D box snoRNA が持つ2つのモチーフ配列に挟まれた配列を *C. intestinalis* ゲノムから抽出し、BPLA カーネルを用いて snoRNA か否かを判定させた。次に snoRNA と判定された配列からプライマーを作成し、5' (3') RACE によって発現確認と配列を決定した。全長を決定した有力候補について再度ホヤゲノムに BLAST を用いて当たるところ、2つの隣接する遺伝子、GDP-L-fucose synthetase 遺伝子と tRNA pseudouridine synthetase 遺伝子のイントロン領域とその遺伝子間の intergenic 領域の合計8箇所にコーディングされてクラスターを形成していることが判明した (図3)。最終的に得られた snoRNA 候補に対して *in situ* hybridization を行うことで各 snoRNA 候補の細胞内局在を決定することに成功した (図4)。

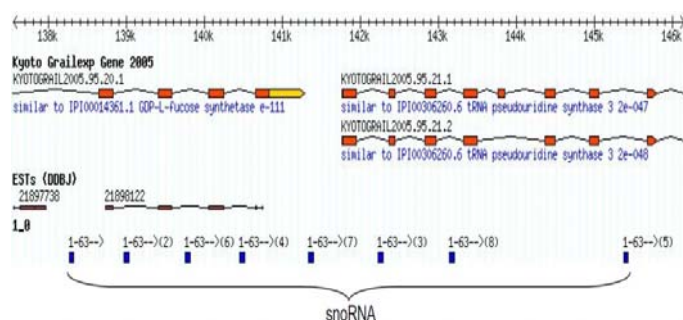


図 3. 8箇所へマップされた snoRNA 候補



図 4. *in situ* hybridization による予測 snoRNA の局在同定

#### (5) BPLA カーネルのメダカゲノムへの適用

これまで開発してきた手法を用いて実際のゲノム上から機能性 RNA コード領域を網羅的に予測し、その予測候補に対して実際に RACE 法による機能性 RNA コード領域の全長決定とシーケンサーによる配列決定を行い、その結果をカーネル関数の学習にフィードバックすることにより、機能性 RNA 予測の情報システムの精度を向上させることを第一の目的とした。この目的を達成する研究として、メダカゲノムにおける機能性 RNA の探索と解析を行った。実験対象生物として、ヒトと類似した遺伝子組成を有するメダカを選択した。BPLA カーネルを用いてゲノムから snoRNA の配列予測を行った。その際、メダカゲノムと同じ二塩基組成を持つ疑似メダカゲノムを作成し、配列予測を行うにあたっての条件検討を行った。次に、メダカゲノム上から C/D box モチーフに挟まれる領域を抽出し、これをテストデータとして予測を行った。さらに、既存の snoRNA 配列予測ツールである snoSCAN を用いて予測を行い、BPLA カーネルによる予測結果との比較を行った (表 3)。インシリコ実験で得た 95 個の予測配列をウェット実験によって検証した。まず、Hd-rR 系統のメダカの産卵後 7 日経過した受精卵(ステージ 39)から total RNA を抽出した。次に予測した配列に対し、10 塩基程度離してプライマーを 2 カ所で設計した。これを用いて 5' RACE 法を行い、各予測配列に対し発現確認を行った (図 5)。ここで発現が確認された配列をクローニングし、

塩基配列を決定した。5' RACE 法により、データベースに登録された snoRNA 配列 15 個、新規 snoRNA 配列 2 個 (図 6) の発現を確認した、図 6 (b):OL-BPLA93 はデータベースに登録されている snoRNA (SNORD60) と共通の特徴を有していることから、同様の機能を持つ snoRNA である可能性が高い(図 7)。また、BPLA カーネルと snoSCAN で共通に予測されたものは 14 個で、ターゲット配列とメチル化部位を推定することができた。

表 3. BPLA カーネルによる予測結果と、BPLA カーネルと snoSCAN の共通予測結果

Length[nt]	40-50	50-60	60-70	70-80
BPLA kernel	11	49	33	2
BPLA×snoSCAN	0	9	11	2
DB snoRNA	0	8	11	0

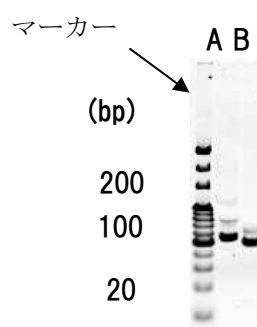


図 5. 5' RACE 法による snoRNA の発現確認: 1 本の予測配列に対し、10 bp 程度ずらしてプライマーを設計した。A, B はその結果増幅された長さの異なる部分配列である (矢印)。

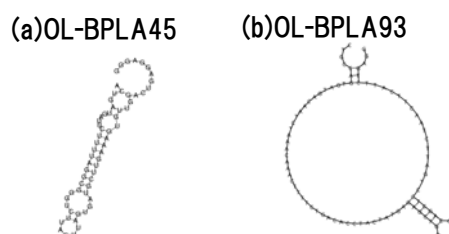


図 6. 本研究で発見された新規 snoRNA

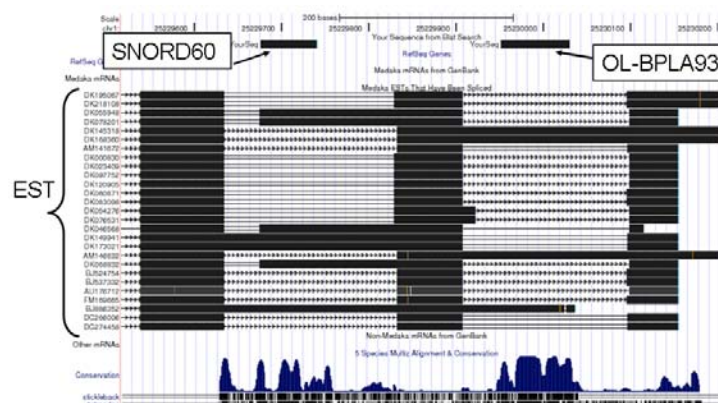


図7. 図6 (b) を BLAT で貼り付けた結果 (下の青いピークは他の魚類との配列保存度を示す)

#### (6) BPLA カーネルのプロファイル拡張

BPLA カーネルのアルゴリズムを拡張し、ゲノムアラインメントからの機能性 RNA の探索に適用できる新しい予測手法を提案した。提案手法は、機能性 RNA の二次構造の情報を、単一の配列ではなくアラインメントデータから抽出することによって、予測の信頼性を向上する。本手法は、アラインメントのベンチマークデータを用いた評価実験において、既存手法のステムカーネルよりも高い予測精度を示した (表4)。また、拡張された BPLA カーネルの予測ツールとしての実用性や性能特性に関して様々な観点から検討し、既存手法との詳細な比較を行なった。その結果、既存手法の RNAz (SCI) やステムカーネルが苦手とする、ノイズの混入した低品質なアラインメントデータに対する予測精度の頑健性 (図8)、未知の機能性 RNA ファミリーの探索 (図9) といった実用上の重要な課題に対して、本手法の有効性が明らかになった。今回提案された拡張 BPLA カーネルは、ゲノムアラインメントからの機能性 RNA の探索を高精度に行なうことができる。本手法は、入力のアラインメントデータがノイズを含む場合でも比較的頑健な予測が可能であり、探索ツールとして実用性が高いと言える。

表 4. BPLA カーネルとステムカーネルの予測精度の比較

Family	ROC score	
	BPLA kernel	Stem kernel
C/D snoRNA	0.951	0.779
H/ACA snoRNA	0.973	0.865
miRNA precursor	0.960	0.880
RNase P RNA	0.913	0.825
SRP RNA	0.998	0.977
Riboswitch	0.946	0.805
tmRNA	0.998	0.946
tRNA	0.987	0.929
Average	0.966	0.876

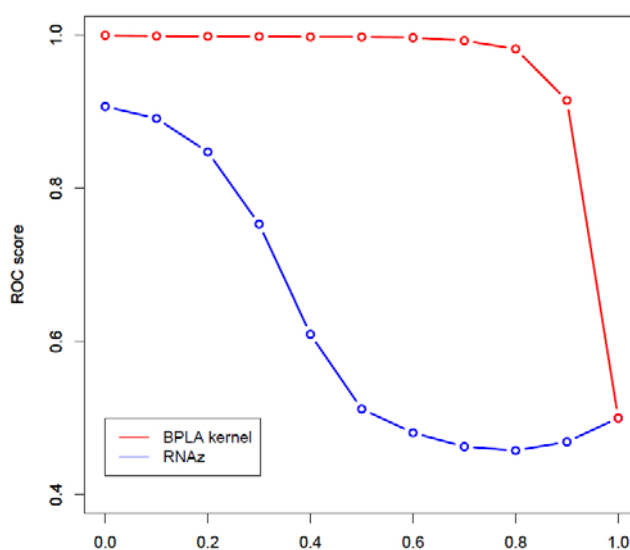


図 8. 低品質なアラインメントデータに対する予測精度。入力のアラインメントデータに含まれるエラーの量に対して、BPLA カーネルと既存手法 RNAz の予測精度の頑健性を評価した。ROC スコアは 1 に近いほど予測精度が高いことを示す。

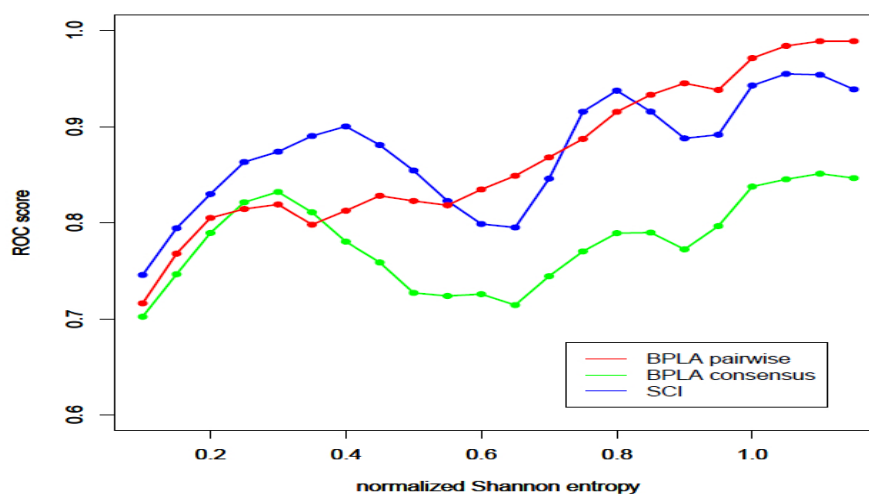


図9. 未知の機能性 RNA ファミリーに対する予測精度の比較。進化的距離の様々なアラインメントについて、提案手法 BPLA カーネルと既存手法 SCI の予測精度を評価した。normalized Shannon entropy は大きいほど進化的距離が遠いことを示し、ROC スコアは 1 に近いほど予測精度が高いことを示す

#### (7) BPLA カーネルのパラメータ最適化法の開発

BPLA カーネルには、いくつかのハイパーパラメータが存在し、その値が性能を大きく左右するため、グリッドサーチなどで最適化する必要があるが、これには多くの時間が必要となる。そこで、勾配法を基にしたハイパーパラメータ最適化法を開発し、表 5 に示すように最適化に要する時間を大きく削減することに成功した (Sato et al., Genome Informatics, 23(1), pp.128-138, 2009)。

表 5. 本手法とグリッドサーチによるハイパーパラメータ最適化の比較

Family	gradient-based optimization			grid search		
	N	AUC	time (h)	N	AUC	time (h)
5S rRNA	24	0.997 (0.956)	1.8	7776	0.997	149.0
tRNA	15	0.993 (0.957)	0.6	7776	0.995	67.5
C/D snoRNA	25	0.843 (0.773)	1.8	7776	0.887	124.8
H/ACA snoRNA	23	0.879 (0.832)	3.5	7776	0.906	248.0
miRNA	20	0.995 (0.971)	1.1	7776	0.995	89.0

## (8) 比較ゲノムによる機能性 RNA の大規模探索手法の開発

これまで開発してきた機能性 RNA 配列解析のためのカーネル関数群を、比較ゲノム手法と組み合わせることにより、任意の複数種のゲノム配列群から機能性 RNA 領域を高速に検出する手法を開発した。RNAz に代表される機能性 RNA 識別手法の多くは、入力として配列アラインメントを必要とする。そのため、機能性 RNA のゲノム探索では、複数種のゲノムを比較して保存領域を抽出する処理が重要となる。しかし、既存の比較ゲノム手法の多くは塩基配列の相同性のみ注目しており、機能性 RNA の特徴である二次構造の保存を考慮していない。そこで、ゲノムの二次構造を高速に比較する新しい手法を提案した。二次構造の文字列表記法を利用し、RNALfold で得られる二次構造予測の結果を文字列として直接比較する (図 10)。比較のアルゴリズムには、当研究グループで開発された極めて高速な比較ゲノムツールである Murasaki を用いる。これにより、数億塩基対からなる高等生物の広大なゲノムから短時間で二次構造の保存された領域を検出することができる。本手法の有効性を示すために、線虫 *C. elegans* のゲノムを同じ線形動物門に属する *C. briggsae* および *P. pacificus* と比較した。RNAz の主要な識別指標である MFE Z-score と SCI を用いた評価により、本手法で検出された領域は従来の比較ゲノム手法よりも熱力学的に安定で有意に保存された 2 次構造を形成しうることが示唆された (図 11)。

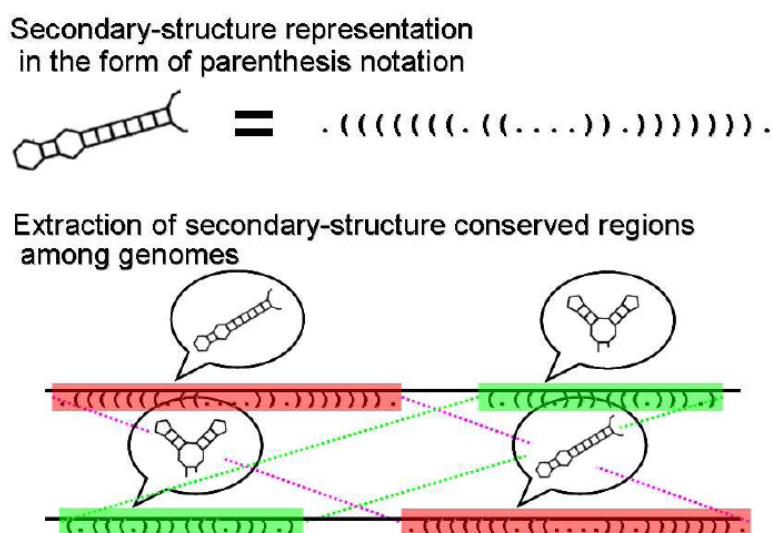


図 10. 本手法の概略：(上段) RNA の 2 次構造の文字列表記法 (下段) ゲノム上に予測された 2 次構造を文字列として比較し、保存領域を高速に検出する。

表 6. 抽出された領域の熱力学安定性による評価

	Our method	homology-based
vs. <i>C. briggsae</i>		
MFE Z-score	-1.52	-0.35
similarity	35.8%	80.6%
coverage	0.43%	0.66%
vs. <i>P. pacificus</i>		
MFE Z-score	-2.27	-0.86
similarity	34.6%	68.6%
coverage	0.035%	0.049%

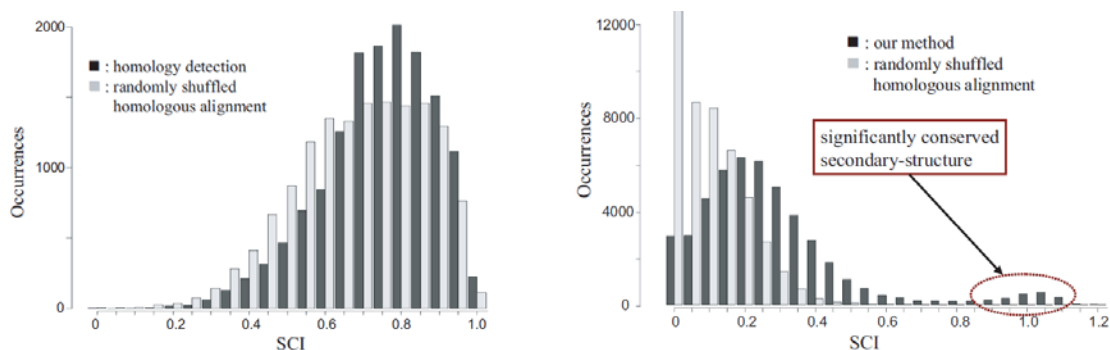


図 1 1. 抽出された領域における 2 次構造の保存. SCI の値を対応する帰無分布と比較した。高い SCI は、その領域に強く保存された 2 次構造が存在することを意味する。左：塩基配列の相同性に注目する従来手法では、SCI の値は帰無分布と差が無かった。右：2 次構造の保存に注目する本手法は、帰無分布よりも有意に SCI の高い領域を検出することに成功した。

## 2. 1. 1. 3. 3. まとめ

本研究課題では、RNA 配列解析のための新しい確率文法モデルを開発した。とくに、ノンパラメトリックベイズ法による確率文法モデルは与えられたデータから最適な文法規則を学習することが可能であり、RNA 二次構造に関する解析（二次構造予測、二次構造を考慮した相同性検索、二次構造を持つ機能性 RNA 遺伝子発見など）に広く応用することが可能である。

また、カーネル法による RNA 配列解析技術を確立し、ウェット実験でその実用性を確認した。とくに、これまでの手法では難しいとされてきた snoRNA の予測において非常に高い精度が得られており、本手法の有効性を示すことができた。カーネル法による解析は、相同性検索の他に、機能性 RNA 遺伝子の発見、クラスタリングによる未知遺伝子の機能解析、mRNA との相関分析などに応用することが可能である。

これらの技術は RNA 配列解析における基盤技術と位置付けられるものであり、これらを基にしてさまざまなアプリケーションの開発が可能である。このことから、機能性 RNA 解析のためのバイオインフォマティクスの中核を担う技術の一つであると言える。



## 2. 1. 2. ゲノム配列からの機能性 RNA の網羅的予測

JBIC 集中研①、分室 2 (インテック W&G)

共同実施先： 産総研①、東京大学 (1 ; 浅井)、三菱総合研究所

ヒトなどの高等生物ではタンパクをコードしていない RNA が数多く存在し、発生・分化や疾患の過程において重要な役割を果たしていることが明らかになってきた。それらは機能性 RNA または non-coding RNA (ncRNA) と呼ばれている。機能性 RNA の生体内での役割を理解するためには、ゲノムに存在する機能性 RNA を網羅的に発見する必要がある。本項目では、ヒトゲノムから機能性 RNA を効率よく発見するインフォマティクス技術確立すること目標として研究を進めた。そして、インフォマティクス技術とマイクロアレイによる発現解析を組み合わせて機能性 RNA を効率よく発見する基盤技術を開発した。また、産業上重要である micro RNA (miRNA) を高精度に発見し機能解析するインフォマティクス技術を開発し、論文発表と特許出願を行った。また、偽遺伝子の特徴を利用した機能性 RNA の予測技術を開発し、その有効性を示した。そして、発見された機能性 RNA 候補の中から癌の診断マーカーとなるものを見出し特許出願を行った。

平成 18 年度はヒトゲノムから機能性 RNA を網羅的に予測するパイプラインを開発した。そして平成 19 年度以降から予測した機能性 RNA に対するマイクロアレイを用いた発現解析実験を行った。また、平成 19 年度以降は miRNA の予測/機能解析手法と、偽遺伝子の特徴を利用した機能性 RNA 予測手法を開発した。プロジェクト期間における研究の進捗状況を図 1 に記す。

		平成				
		17	18	19	20	21
機能性 RNA 予測パイプラインの開発			→			
アレイによる発現確認				→	→	→
独自手法の開発	miRNA 予測		→	→	→	→
	偽遺伝子の特徴に基づく予測				→	→

図 1. プロジェクト期間における研究の進捗

### 2. 1. 2. 1. 機能性 RNA 予測パイプラインの開発

#### (1) 概要

RNAz や snoSeeker などの既存の機能性 RNA 予測手法と、我々のグループで独自開発したアルゴリズムを組み合わせて機能性 RNA の網羅的な予測を行うパイプラインを開発した。

我々のパイプラインは、①オルソログスな領域からの機能性 RNA 予測、②パラログスな領域からの機能性 RNA 予測、③2次構造モチーフに基づく機能性 RNA 予測の3種類かからなる。以下にそれぞれのパイプラインの概要を記す。

#### ①オルソログスな領域からの機能性 RNA 予測 (Synseg パイプライン)

多くの機能性 RNA は重要な機能をもち進化的に保存されている。したがって、機能性 RNA の予測には生物種間の比較ゲノムが有効である。我々はシンテニー領域（遺伝子の並び順が保存された領域）にある遺伝子間領域を独自開発した mxscarna プログラムにより網羅的にアライメントした。mxscarna は2次構造を考慮したアライメントを行うため、機能性 RNA 領域を正確にアライメントすることができる。また、非常に高速であるため大規模な領域に対してアライメントを実行できる。得られたアライメントに対して RNAz プログラムを実行することで機能性 RNA を予測する。

#### ②パラログスな領域からの機能性 RNA 予測 (Self-Chain パイプライン)

多くの機能性 RNA の多くはヒトゲノム内にパラログとして存在する。実際、我々はヒトゲノム同士の相同性検索により得られた自己アライメントのなかに既知機能性 RNA の約80%が含まれていることを見出した。本パイプラインでは、まずヒトゲノム同士の相同性検索を CBRC で開発された Last プログラム (<http://last.cbrc.jp/>) により行い、得られた自己アライメントに対して RNAz プログラムと snoSeeker プログラムを実行することにより機能性 RNA を予測する。本パイプラインでは進化的に保存度の高くない機能性 RNA を検出できる可能性がある。

#### ③2次構造モチーフに基づく機能性 RNA 予測 (RNAmine パイプライン)

我々が開発した RNAmine プログラムは大量の DNA 配列が与えられたときに2次構造モチーフを抽出するツールである。本パイプラインではヒトゲノム内で進化的に保存した領域を大量に収集し、RNAmine プログラムを実行することで2次構造モチーフを抽出する。そして抽出した2次構造モチーフをクエリーとして infernal プログラム (<http://infernal.janelia.org/>) を使ってヒトゲノムを検索することで機能性 RNA を予測する。

### (2) マイクロアレイ実験

上記①～③のパイプラインに加え、miRRim プログラム（独自開発）による miRNA 候補、進化的に保存度の高い遺伝子間領域 (Ultraconserved region)、独自のコントロール領域などの加え、合計 10000 以上の機能性 RNA 候補に対してマイクロアレイ実験を実施した。表 1 はマイクロアレイに載せた機能性 RNA 候補の内訳である。

マイクロアレイは実験コストに優れたアジレント・テクノロジー社の 44K フォーマット

アレイ（1枚のアレイに41000以上のプローブを搭載可能）を用いた。さらにプローブ設計を自前で行うことで実験に必要なコストを下げる事が出来た。実験に必要な total RNA は実験グループの廣瀬研究室からいただいた。アレイ実験は、日立ソフトエンジニアリング(株)の横井氏の協力により行われた。

表1. マイクロアレイに載せた領域の内訳

種別	候補数	説明	高発現 <sup>a)</sup>
①SelfChainパイプライン	2158	パラログ領域の比較ゲノムに基づく機能性RNA候補	267
②Synsegパイプライン	5310	オルソログ領域の比較ゲノムに基づく機能性RNA候補	627
③RNAmineパイプライン	108	RNAmineによる2次構造モチーフ	11
miRNA候補	279	miRRimによるmiRNA予測	24
Ultra conserved region	1325	進化的に強く保存された領域	108
Ultra stable fold	30	非常に安定した2次構造を持つ領域	3
Human accerelated region	49	ヒトだけで置換が見られる領域	2
Inverted duplication	29	逆位重複に由来するヘアピン構造	1
コントロール1	468	Rfamに登録された既知機能性RNA	262
コントロール2	211	既知のタンパクコード遺伝子	100

a) 4つ以上の組織/細胞で発現量が高い領域

#### (2) - 1. プローブ設計

プローブの設計はフリーソフトである OligoArray2.0

(<http://berry.engin.umich.edu/oligoarray2/>)を用いて行った。既知のタンパクコード遺伝子を第1の背景配列として用い、これらにクロスハイブリするとみなされたプローブはすべて除外した。さらにヒトゲノム全域を第2の背景配列として用いた。ヒトゲノムへのクロスハイブリの個数はプローブ選択の基準のひとつとして用いられた。

#### (2) - 2. 実験した組織/細胞

平成19年度中に Brain、Testis、Hela cell、Thymus の4組織/細胞に対する予備的な実験結果を得た。tRNA や snoRNA など様々な組織で発現量が高い機能性RNAについては、高い発現強度が検出できた。このことは我々の設計したアレイの有効性を示すものである。そこで、平成20年度は、さらに癌細胞、癌細胞と隣接する正常細胞、核や核小体から抽出したRNAなどに対するマイクロアレイ実験を行い、合計で23組織/細胞における網羅的な発現データを得た。表2に、23組織/細胞の内訳を記す。

表 2. マイクロアレイ実験を行った 23 組織/細胞

組織/細胞	画分
1 HeLa	
2 Brain	
3 Thymus	
4 Liver	
5 Lung	
6 Muscle	
7 Prostate	
8 Testis	
9 Bladder	Normal
10 Bladder	Tumor
11 Kidney	
12 Spleen	
13 Cervix	
14 Colon	Normal
15 Colon	Tumor
16 Liver	Normal
17 Liver	Tumor
18 Prostate	Normal
19 Prostate	Tumor
20 HeLa	Cytosplasm
21 HeLa	Total
22 HeLa	Nucleolar
23 HeLa	Nuclear

### (3) マイクロアレイ結果に基づく有望な機能性 RNA 候補の抽出

マイクロアレイはノイズの多い実験手法なので、有望な機能性 RNA 候補を抽出する際には複数の実験で発現量が高いものを選択すべきである。そこで、4 つ以上の組織/細胞で基準値以上発現している候補を“高発現”と判定した。一般に snoRNA は発現量が十分に高いことが知られている。そこで、コントロールとしてマイクロアレイに載せた snoRNA の発現量を基準値として用いた。表 1 の第 4 カラムにグループごとの高発現領域数を記す。コントロールとして用いた既知の機能性 RNA はやはり高発現と判定される割合が多かった。

### (4) Exosome のノックダウン実験

Exosome は細胞内の様々なタイプの RNA 分子を分解するタンパク質複合体である。Exosome をノックアウトすれば普段は分解されて発現が観測できない RNA の発現を観測できると考えられる。図 2 は Exosome ノックダウン前後におけるマイクロアレイ結果の比較である。ノックダウン実験は HeLa 細胞を用いて行った (機能解析グループ 廣瀬研究室)。対角線よりも右側にあるプロットは、Exosome ノックダウン後に発現量が増加したプローブである。Exosome ノックダウン実験により発現量が増加したプローブを 243 個抽出した (図 2 の点線丸の部分)。これらのプローブから 15 個を選択しリアルタイム PCR による検証実験を行ったところ 10 個が Exosome ノックダウン後に発現量が増加することが示された。したがって、

Exosome ノックダウン実験により効率よく発現している機能性 RNA を発見できると考えられる。また、Exosome ノックダウン実験により発現量が増加したプローブ（243 個）と表 1 で高発現と判定された候補を比較すると、重複は 17 個のみであった。したがって、Exosome の実験により必ずしも発現量が高くないが発現している機能性 RNA 候補を効率よく発見できると考えられる。

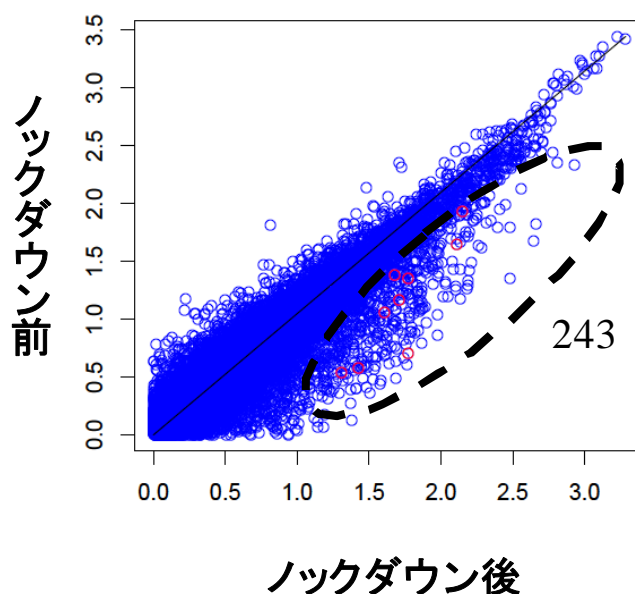


図 2. Exosome ノックダウン後の発現量の変化。

X 軸、Y 軸の数値はノーマライズした発現強度である。

#### (5) まとめ

インフォマティクス技術により 1 万個の機能性 RNA 候補を抽出し、マイクロアレイを利用することで合計 1,500 個以上の発現している機能性 RNA 候補を抽出することが出来た。Exosome ノックダウン実験では、発現量が必ずしも高くない機能性 RNA を効率よく発見出来るという点で独自性がある。我々が行った機能性 RNA 候補の抽出、すなわちインフォマティクス技術とマイクロアレイを組み合わせたパイプラインは、発現している機能性 RNA を効率よく発見するために有効な基盤技術であると考えられる。現在、本パイプラインで発見した機能性 RNA の中から生物学的に興味深いものを選択し詳細な実験を行っている。一例として、機能解析グループ廣瀬研究室では我々が発見した機能性 RNA 候補の中からエピジェネティックな遺伝子制御に関与するであろう候補を発見し機能解析を進めている。

## 2. 1. 2. 2. miRNA 遺伝子の予測と機能解析手法の開発

### (1) 概要

miRNA はその部分塩基配列が標的であるmRNAに配列相補的に結合し、遺伝子発現制御に関与する重要な非コードRNA のファミリーである。ヒトゲノムに見つかるmiRNAは年々増え続けており未発見のmiRNAが多数存在すると考えられる。我々は平成19年度からmiRNA予測手法の開発に着手し、高精度miRNA予測プログラムmiRRimを開発した (Terai et al., RNA 13(12), pp. 2081-2090, 2007)。miRRimで予測されたmiRNAは、機能解析グループ塩見研究室のago-associated small RNAの検出結果と比較され、いくつかは真のmiRNAであることが示された。その後、更にアルゴリズムを改良し、予測精度の向上と成熟miRNA部位の予測を可能とする手法を開発した(miRRim2)。

### (2) アルゴリズムの概要

miRNA 前駆体は安定したヘアピン構造をとり、その塩基配列は進化的に強く保存されているものが多い。また、miRNA 前駆体ヘアピンのステム領域はループ領域と比較し相対的に強く保存されている傾向がある。miRRim はそれらの傾向を隠れマルコフモデルによりモデル化することでmiRNA を予測する手法である。

miRNA前駆体ヘアピンはDroshaとDicerによるプロセッシングを経て成熟miRNAとなる。最近、miRNA前駆体のプロセッシングに必要な2次構造制約がいくつかのグループにより報告された。たとえばDroshaはmiRNA前駆体のもっとも外側の塩基対を認識し、その12~13bp先の塩基対を切断する。したがって、成熟miRNAの5' 末端の上流12~13bpは塩基対を形成しやすく、そのさらに上流は塩基対を形成しにくいはずである。それら新たに判明した2次構造的特徴を正確に捉えることが出来ればmiRNAの予測精度を向上させることが出来るはずである。現在開発中のmiRRim2ではHMMアーキテクチャをより精緻なものにすることで新たに判明した2次構造的特徴を捉えられるようにした。また、従来のmiRRimでは用いられなかった新たな特徴量を3種類加えることで予測精度を向上させることに成功した。

#### (2) - 1. HMM アーキテクチャ

miRRim2 で用いられている miRNA モデルの HMM アーキテクチャを図 3 に記す。miRNA 前駆体ヘアピンは成熟 miRNA 部位が 5' 側のステムに存在する場合、3' 側のステムに存在する場合、そして両側のステムに存在する場合がある。図 3 で示したアーキテクチャはそれら 3 つのケースに対応したサブアーキテクチャをもっている。これにより、成熟 miRNA 部位に見られる進化的保存度や 2 次構造的特徴を捉えることが出来ると同時に、成熟 miRNA 部位の予測が可能となる。

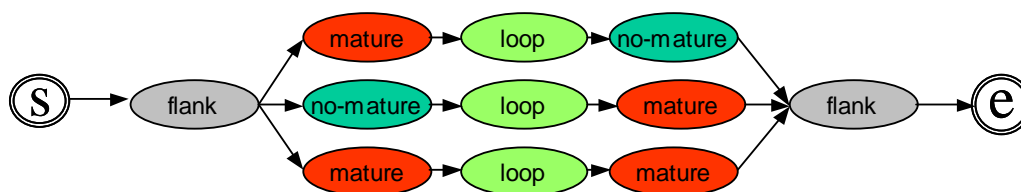


図3. miRNA モデルのアーキテクチャ概要

二重丸で囲った s と e はそれぞれ開始状態と終了状態を表す。楕円はサブアーキテクチャを表す。

### (2) - 2. miRRim2 で用いられる特徴量

miRRim2 ではヒトゲノムの各ポジションを 7 種類の特徴量で表現する。miRRim2 で用いられる 7 つの特徴量は下記のとおりである(下線は新たに追加した特長量である)。

- 特徴量 1. PhastCons の保存度スコア
- 特徴量 2. PhyloP の保存度スコア
- 特徴量 3. 最小自由エネルギーの Z-score (※1)
- 特徴量 4. 最大塩基対確率 (※2)
- 特徴量 5. 塩基対間距離 (※3)
- 特徴量 6. ステム保存度差 (※4)
- 特徴量 7. 塩基 (A, T, G, C)

※1 ある部位を中心として周辺 100-bp 領域の最小自由エネルギーの Z-score

※2 ある部位が他の部位と塩基対を組む確率(塩基対確率)の最大値。

※3 最大塩基対確率が 50%以上のとき、塩基対を組む 2 つの塩基の距離

※4 最大塩基対確率が 50%以上のとき、塩基対を組む 2 つの塩基周辺の保存度差

### (3) 予測精度

図4に miRNA 前駆体ヘアピンの予測精度と他種法との比較を示す。miRRim2 は改良前の miRRim と比較して高精度に miRNA を予測出来る。他種法と比較しても miRRim2 の予測精度は優れている。

miRRim2 の有用な特徴は成熟 miRNA 部位を予測することが出来ることである。成熟 miRNA とターゲット遺伝子の相互作用を規定するのは成熟 miRNA 部位の 5' 末にある seed 配列であることが知られている。したがって miRNA のターゲット予測を行うためには成熟 miRNA の 5' 端を予測することが重要である。図5は、成熟 miRNA 部位の予測精度を示したものである。miRRim2 では Sensitivity≒0.3、PPV≒0.3 で成熟 miRNA の 5' 末端を正確に(1bp の誤りもなく)予測することが出来る。5' 末端の位置を 1bp 隣まで正解と考えると、Sensitivity≒0.5、PPV≒0.5 で予測することが可能である。

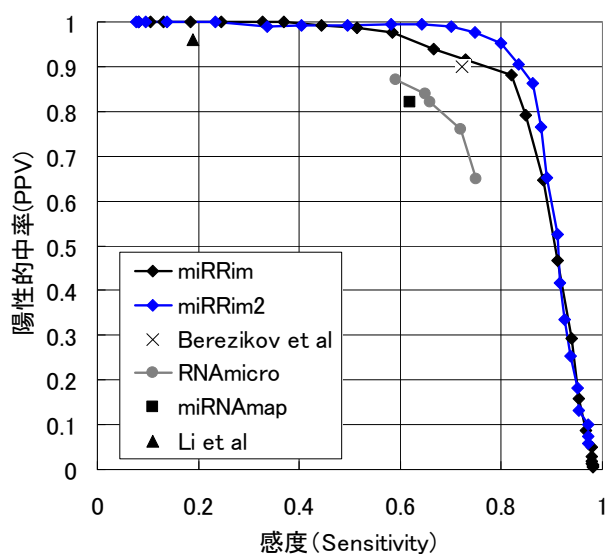


図 4. miRNA 前駆体ヘアピンの予測精度

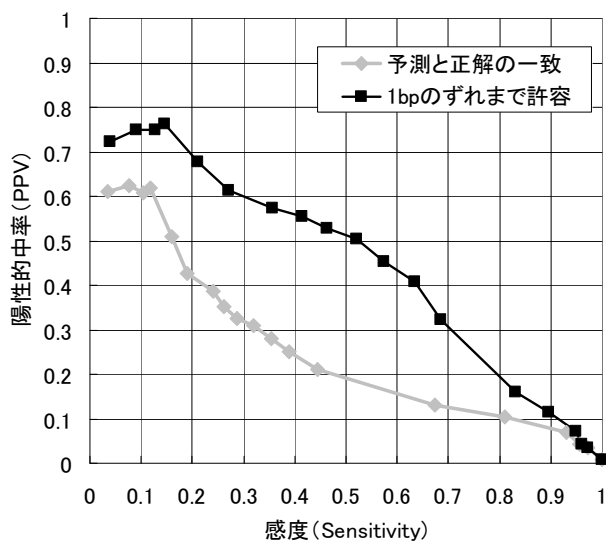


図 5. 成熟 miRNA の 5' 末端の予測精度

(4) まとめ

miRRim2ではmiRNAモデルを精緻化し、新しい特徴量を取り入れることで予測精度を向上させることが出来た。さらにmiRRim2では成熟miRNA部位の予測が可能である。これにより予測したmiRNA前駆体の機能解析 (=ターゲット予測) を行うことが可能である。



## 2. 1. 2. 3. 偽遺伝子の特徴を利用した機能性 RNA 予測手法の開発

### (1) 概要

見つけたい遺伝子の偽遺伝子を見つけることが出来れば、その相同領域をゲノムから抽出することで目的の遺伝子を発見することが出来る。問題は目的の遺伝子を知ることなしに、どのように目的遺伝子の偽遺伝子だけを発見するかである。

レトロ転移により生じた偽遺伝子には2タイプの特徴が生じる。1つは偽遺伝子周辺に現れる Poly-A や、Target Site Duplication (TSD) と呼ばれる短い重複配列である。もう1つは、偽遺伝子にしばしば見られる 5' 側の逆位 (5' -inversion) である。我々はそれらの特徴の有無に基づき偽遺伝子とその親遺伝子のペアを発見する手法を開発した。発見した偽遺伝子：親遺伝子ペアのなかからタンパクコード遺伝子を除くことで機能性 RNA の偽遺伝子：親遺伝子ペアを抽出する。

### (2) 偽遺伝子周辺の特徴を利用した偽遺伝子：親遺伝子ペアの抽出

レトロ転移により生じた偽遺伝子の周辺には3種類の配列特徴が生じることが知られている (図6)。1つめは偽遺伝子の3' 側に生じる長さ5~20bp程度の Poly-A tail である。2つ目は、偽遺伝子の両側に生じる Target site duplication (TSD) と呼ばれる5~20bp程度の重複配列である。3つ目は5' 側の TSD と重複して現れる TTAAAA モチーフである。この TTAAAA モチーフを持たない偽遺伝子もしばしば存在する。



図6. 偽遺伝子周辺の配列特徴

我々は上記の3つの配列特徴を高精度に検出するアルゴリズムを開発し、ソフトウェア TSDscan として実装した。TSDscan では、偽遺伝子候補配列の上流と下配列を特殊なスコアリングによりアライメントすることで Poly-A tail や TSD を検出する。

TSDscan を使って図7に示すような領域をヒトゲノムから抽出する。まず、ヒトゲノム内で相同性のある領域ペアを網羅的に同定する。ある領域ペアのどちらかの領域に Poly-A tail-TSD が検出された場合、その領域ペアを偽遺伝子：親遺伝子ペアとみなす。最後にタンパクコード遺伝子に相同性のある偽遺伝子：親遺伝子ペアを除く。残った領域を機能性 RNA の偽遺伝子：親偽遺伝子ペア候補とする。



図7. 遺伝子・偽遺伝子領域ペアの配列パターン

## (3) 5' -inversion を利用した遺伝子・偽遺伝子ペアの抽出

遺伝子から転写された mRNA が逆転写されゲノムに挿入される際には、その mRNA の 5' 側が逆位 (5' -inversion) を起こすことがある。そこで、図 8 のような相同性パターンをもつ領域ペアをヒトゲノムから網羅的に抽出する。そして、タンパクコード領域に相同性のある領域ペアを取り除くことで機能性 RNA の偽遺伝子：親遺伝子ペアを抽出する。

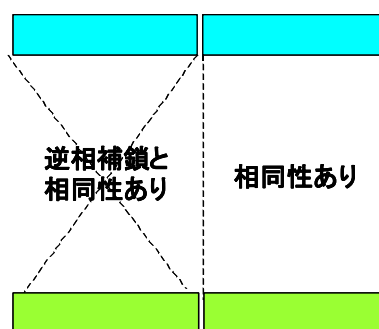


図8. 5' -inversion をもつ偽遺伝子：親遺伝子領域ペアの配列パターン

## (4) 偽遺伝子：親遺伝子ペアの抽出結果

機能性 RNA に由来する偽遺伝子：親遺伝子ペアの抽出結果を図 9 に記す。偽遺伝子周辺の特徴を利用して 280822 個のペアを検出した。タンパク質遺伝子に由来すると思われるものを除くことで 1644 個のペアを得た (①)。5' -inversion の特徴を利用して 660 個のペアを検出した。タンパク質遺伝子に由来すると思われるものを除くことで 29 個のペアを得た (②)。①と②が重複するのは 1 ペアであった。

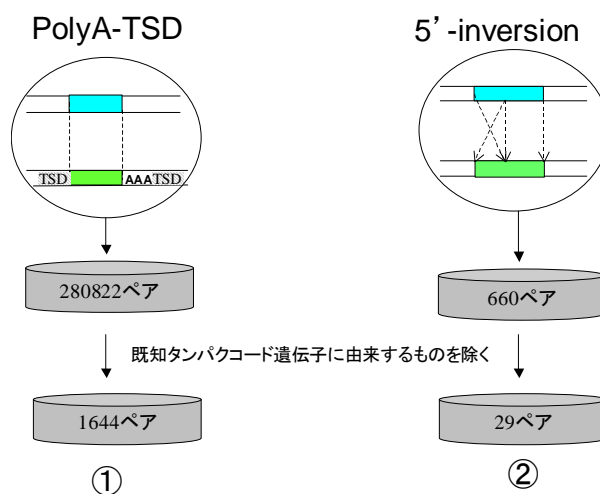


図9．機能性 RNA 由来の偽遺伝子：遺伝子領域ペアの抽出結果

#### (5) 発見した新規機能性 RNA 候補の一例

上記手法で発見された機能性 RNA の一例を図 10 に記す。この機能性 RNA (以後、R75) はタンパクコード遺伝子のイントロンにあり 2 つの偽遺伝子を持っている (図 10)。R75 の予測 2 次構造は H/ACA 型 snoRNA の典型である 2 つの長いステムを持ち、ループ領域には進化的に保存された ACA-ボックスを持っている。このことから、R75 は新規の H/ACA 型 snoRNA 遺伝子であると考えられる。機能解析グループ廣瀬研究室の協力により R75 に対して Ribonuclease protection assay を実施したところ、R75 は確かに発現していることを確認した。

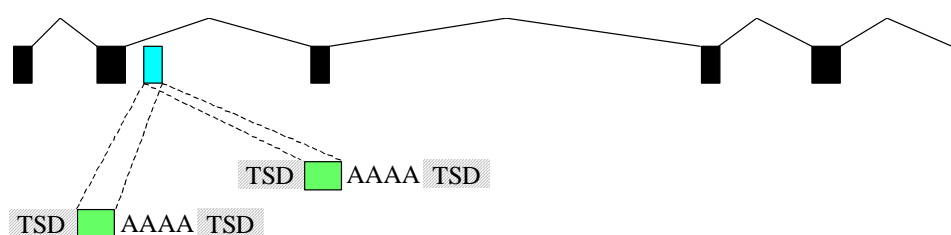


図 10．発見した機能性 RNA 候補の一例

水色のボックスは、機能性 RNA 候補 (R75)、緑色のボックスは R17 の偽遺伝子、黒のボックスはタンパクコード遺伝子のエクソンを示す。

図 11(a) は R75 の 2 次構造を模式的に書いたものである。ほとんどの H/ACA 型 snoRNA は図 11(b) のような構造をとるが、R75 は典型的な snoRNA とは違う構造をとるようである。

偽遺伝子の特徴を利用した方法は、既知機能性 RNA との配列/2次構造の類似性に依存しないため、このような非典型的な機能性 RNA を発見できるという強みをもつ。

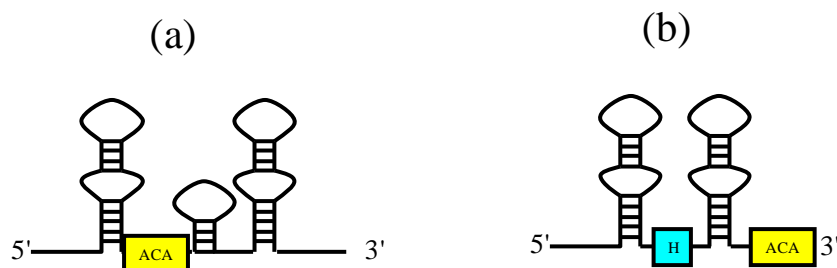


図 1.1. R75 の 2 次構造の模式図

(a) R75 の予測 2 次構造、(b) 典型的な H/ACA snoRNA の 2 次構造

#### (6) まとめ

偽遺伝子の特徴を利用した機能性 RNA 発見手法を提案し、ヒトゲノムに対して網羅的に実行した。予測結果の中には新規 H/ACA 型 snoRNA と思われる候補や、がん細胞特異的に発現する機能性 RNA 候補（結果は示さない）が含まれていることを発見した。本手法は既存手法では用いられない特徴を予測に用いるので、これまで発見できなかった機能性 RNA を発見できる可能性がある。TSDscan については、Terai et al., *Nucleic Acids Research* 38(4), pp. 1163–1171, 2010 に論文発表した。

### 2. 1. 2. 4. 結論

本研究課題ではヒトゲノムから 1 万個以上の機能性 RNA をインフォマティクス技術により予測し、カスタムアレイを使って 1500 個以上の発現している機能性 RNA 候補を抽出した。このプロセスは機能性 RNA を効率よく発見する基盤として有用である。また産業上重要である miRNA 遺伝子を現時点では世界最高精度で予測できる手法を開発した(miRRim2)。miRRim2 では、成熟 miRNA 部位を予測することが出来るので、予測した miRNA の機能推定が可能である。また偽遺伝子の特徴を利用した機能性 RNA 発見手法を提案した。有望な候補について実験的検証を行い、提案手法の有用性を示した。さらに提案手法で予測した機能性 RNA 候補のなかから癌特異的に発現するものを発見し、癌診断マーカーとして特許出願を行った。

### 2. 1. 3. 機能性RNAデータベースの構築

JBIC 集中研①、分室 2 (インテック W&G)

共同実施先： 産総研①、東京大学 (1 ; 浅井)

従来のほとんどのゲノム解析プロジェクトやゲノムデータベースが、タンパク質コード遺伝子を中心に構築されてきたために、機能性 RNA に特化し、かつ網羅的なデータベースは不十分であった。本研究開発項目では、単に既知の機能性 RNA 配列のデータを蓄積するだけでなく、予測した機能性 RNA 配列、予測の元となった様々な情報解析結果、次世代シーケンサーによって読まれた機能性 RNA 断片のデータなどを集積し、新規機能性 RNA の発見と絞り込み、機能解析に活用できるデータベースの開発を行うことを目標に、研究開発を進めることにした。同時に、実験グループから提供される RNA 配列データを解析し、開発したデータベースを活用して新規機能性 RNA の発見する研究にも取り組んだ。

#### 2. 1. 3. 1. 機能性 RNA データベースの設計と構築

既知の機能性 RNA とともに、新たに予測した新規機能性 RNA 候補について、塩基配列情報、ゲノム上の位置情報、組織別発現情報、文献情報などの各種情報を統合した機能性 RNA データベース (fRNAdb) を設計した。さらに、fRNAdb と連携して動作するゲノムブラウザ UCSC GenomeBrowser for Functional RNAs を開発して、ゲノム配列上における機能性 RNA の領域を可視化し、プロジェクト内で利用可能としたほか、一部を一般公開し、Kin et al., Nucleic Acids Research 35(Database issue), pp.D145-D148, 2007 で論文発表した。

機能性 RNA データベースは、図 1 に示すような、バイオインフォマティクスグループ内での綿密な連携によって構築された。

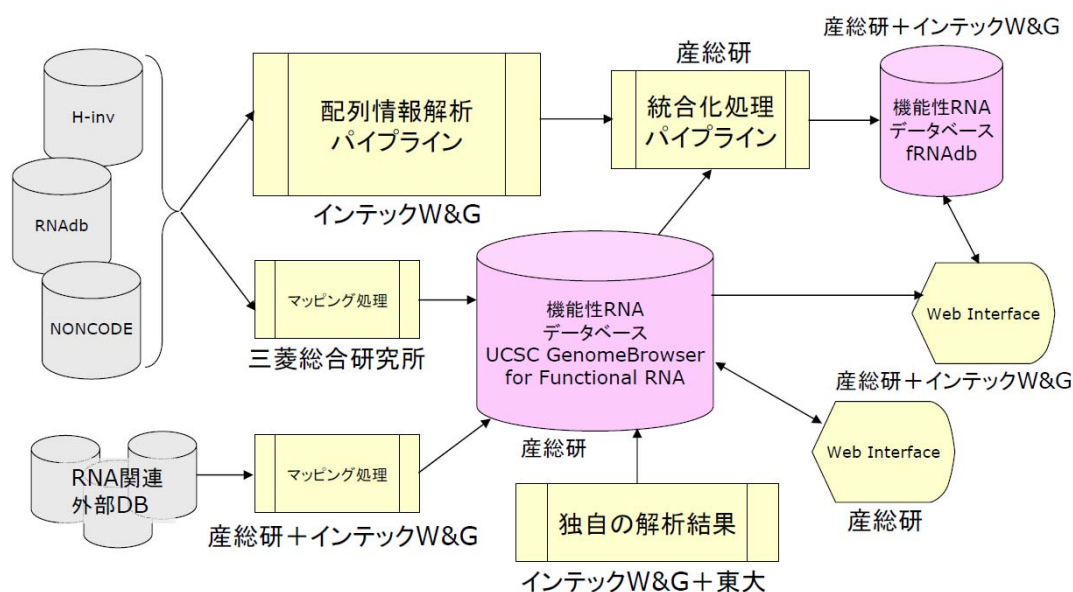


図 1. バイオインフォマティクスグループ内での綿密な連携によるデータベースの構築

機能性 RNA データベースの最も重要な用途は、既知の機能性 RNA と本プロジェクトで予測された新規機能性 RNA 候補情報の閲覧・検索である。図 2 に典型的な検索例を示す。各配列に関連付いた情報は詳細画面にて、図 3 に示すメッセンジャービジュアルライザーによって視覚化される。

### fRNAdb --- Main Listing

This is the top page of the functional RNA database. Please click "refresh" button to perform query.

Last Update of Main Listing Table (summary): 2007/05/29 17:20:58 GMT+0

[View Table Scheme with Statistics](#) or [Class Summary](#) to get the overview of current dataset. Or go to [Help](#) folder.

#### Repository

Load/Save your selection condition from/to your personal folder

Select an item to load    View your data folder

Class Basic DB/ID Expert Sort Column

#### Class Selection

Class selection lets you choose the target RNA class. Please use **Ctrl+Left Mouse Button**(for Windows or Unix) or **Apple+Mouse Button**(for Apple) to select multiple items in the list.

group I intron  
group II intron  
H19  
hypothetical short protein  
inhibitor of Rep2 protein synthesis  
IPW  
Literature curated  
male hypermethylated (MHM) region noncoding nuclear RNA  
miRNA  
Mic F RNA  
miRNA  
**miRNA precursor**  
mRNA-like RNA  
noncoding nuclear RNA  
Noncoding transcript  
NTT  
OxyS RNA  
pec RNA  
pRNA  
Ribonuclease MRP RNA  
Ribonuclease P RNA  
Ribozyme  
Sasaki & Hirose 2007  
scaRNA  
...DNA

①閲覧したい RNA  
の種類を選択する

②refresh ボタン  
をクリックする

Display Options

20 items/page

Page:  1 / 32

Total hits: 621

Download

Clicking  you sort table by that column instantly but limited within the showing table. For sort  opens a sequence page, and  opens a Genome Browser window. Legend:  opens a detail information page,  opens a sequence page, and  opens a Genome Browser window.

no	ID	Acc.	Original	Common Name	Class	Length (nt)	# of exons	# of ESTs	# of mapped	GC%	Max. ORF length (aa)	Repeat Coverage (%)	Pin
1	FR006625	Z70203	u1136	let-7 RNA	miRNA precursor	73	0	0	1	42	21	0	
2	FR007360	AF480538	u1871	mir-100	miRNA precursor	80	1	0	1	46	25	0	
3	FR007361	AF480540	u1872	mir-101	miRNA precursor	75	1	0	1	40	18	0	

図 2. 機能性 RNA データベース fRNAdb の画面の一部。マイクロ RNA 前駆体を選択し、配列一覧を表示させている。検索結果一覧画面から、個々の配列の詳細画面や、ゲノムブラウザーへジャンプして、その遺伝子周辺のゲノム情報を閲覧することも可能になっている。

機能性プロジェクト内部では、機能解析グループにて新規機能性 RNA 遺伝子候補絞り込み作業で頻繁に用いられ（産総研廣瀬グループ、東工大相澤グループ、徳島大塩見グループ、理研清澤グループ）、ツール開発グループ（東大鈴木グループ）においても利用された。デ

データベースは2006年10月からインターネット上で公開しているが

(<http://www.ncrna.org/>)、プロジェクト外では米国、ニュージーランド、フランス、ドイツ、デンマークなどからも多数アクセスされている。

配列構造情報ブラウザ fRNAdbは以下のような内容となっている：

- ア) 以下の配列情報を機能性RNA遺伝子候補として、(1) H-invitational の完全長cDNA配列情報でhypothetical short protein coding、non-protein coding と分類されているもの(2) RNAdbの予測RNA遺伝子、を登録した。
- イ) (1) NONCODEデータベース登録配列、(2) RNAdbの既知RNA遺伝子、を配列情報を既知RNA遺伝子として、登録した。
- ウ) 機能性RNAに関連する配列特性30項目(表1)の指標を計算機によって求めた。
- エ) 資料に基づいて配列の検索や絞込みをするためのインターフェースを開発した。

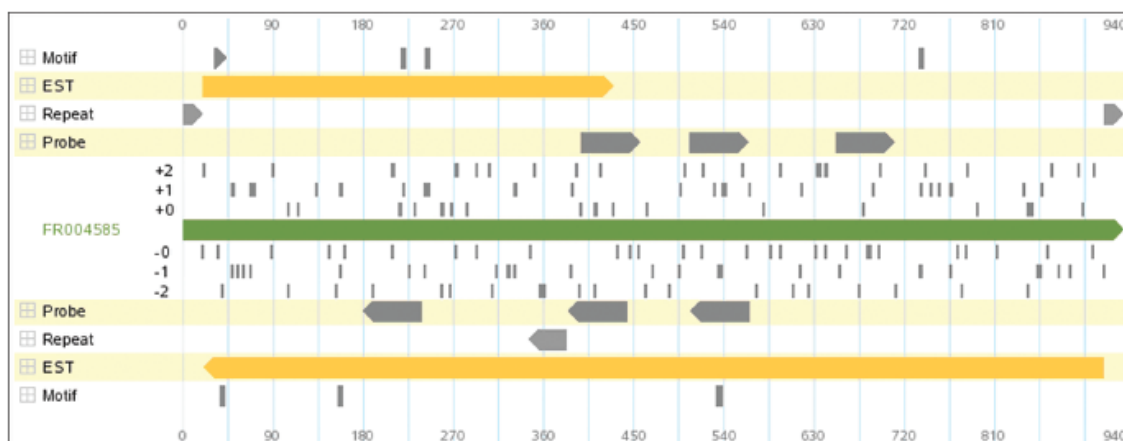


図3. メッセンジャービジュアルライザー画面 配列と EST やモチーフの関係を視覚化する。中央の緑の水平線が cDNA。センス、アンチセンス鎖それぞれのコドンフレームにおける終止コドンの位置と、マイクロアレイ・プローブの位置、リピート配列の位置、EST 配列の位置、Poly(A) シグナルなど、様々な配列モチーフの位置が示されている。

表 1. 機能性RNAに関連する 30 項目の配列特性

Description	
1	Length of the sequence (nt)
2	Number of exons
3	Number of overlapping ESTs
4	Number of mapped positions
5	GC Content (%)
6	Maximum length of potential ORF (aa)
7	Percentage of bases that is covered with repeat elements
8	Repeat elements reside proximal up/down stream
9	Known gene that is a potential sense/antisense of this transcript (exon overlapping)
10	Number of protein homologs (GenBank NR)
11	Known gene that includes this transcript within its intron
12	Known gene region that overlaps with the mapping extent of this transcript (strand not considered)
13	Known gene that overlaps with this transcript within its intron in different strand
14	Known gene where this transcript is possibly a part of its 3UTR
15	Known gene where this transcript is possibly a part of its 5UTR
16	Known gene within upstream 5kbp
17	Known gene within downstream 5kbp
18	Average conservation score over the mapped exonic region
19	Maximum conservation score over the mapped exonic region
20	Maximum conservation score within 500 base upstream from the mapped 5'
21	Overlapping UCSC Ultra Conserved region
22	Number of canonical splice signals in this transcript
23	Number of poly-A signals in this transcript
24	Number of CpG island
25	Associated Transposon Free Region
26	Number of RFAM known RNA motifs in this transcript
27	Number of RNAz predictive RNA motifs in this transcript
28	Number of EvoFold predictive RNA motifs in this transcript
29	Maximum Z-score of RNA secondary structure over this transcript. Scores lower than -6 are significant. Higher scores are considered insignificant. Stored
30	Number of cell lines responding to Affy probes in exon regions of this transcript (Affymetrix Transcriptome Phase 2 Tiling Array Analyses)

ゲノムブラウザ UCSC GenomeBrowser for Functional RNAs には、UCSC Genome Browser をベースに、表 2、表 3 に示す機能性 RNA に特化した情報を追加した。



表 2. 機能性RNAに特化したトラック情報

Track	Description
RNAz folds (6)	Secondary structure annotation of RNAz
ENOR (7)	ENOR (Expressed Noncoding Region) [Lifted from mm5]
Erdmann (8)	Erdmann noncoding RNAs
NONCODE (2)	Mapping information of NONCODE RNAs
RNAdb (3)	Mapping information of RNAdb RNAs
RNA Clusters	Small RNA genes often reside close to each other forming clusters. This track represents computationally identified RNA clusters in human genome.
Rfam seed folds	Genomic search results with INFERNAL and covariance models generated from RFAM seeds.
Rfam full	BLAT mapping results for RFAM full sequence dataset
antisense ChenJ NAR2004 (9)	Sense-antisense pairs among UCSC known genes
tRNAscan-SE (10)	tRNA genes predicted by tRNAscan-SE
Ultra Conserved Elements (11)	100% conserved elements (>=200bp) in human, rat, and mouse
Ultra Conserved Elements 17way	100% conserved elements in 17 vertebrates (longer than 50 bp)
Transposon Free Region (12)	Regions longer than 5Kbp or 10 Kbp containing no LINES, SINEs and LTRs.
Human Accelerated Region (20)	HAR non-coding gene candidates predicted by (20)
Z-score	Regions with Z-score lower (lower is better) than -6 (actual track score=Z-score x

表 3. マイクロRNAに特化したトラック情報

Track	Description
Known miRNAs	miRBase known miRNAs
Predicted miRNAs	miRNAMap and Berezikov's predicted miRNAs
Known targets	TarBase experimentally verified miRNA target sites
Predicted targets	RNAhybrid, PicTar, miRBase, and T-ScanS predicted miRNA

また、機能性 RNA の発見、機能解析を推進し、各ユーザに固有の非公開データと公開データを同時に表示した解析を行うため、以下の機能を追加した。

- ・ 共通2次構造表示機能～比較ゲノムの結果から共通RNA2次構造を予測した結果を視覚的に表示する機能を加えた。
- ・ ユーザー認証機能～ユーザーをユーザー名とパスワードで認証し、ユーザー専用のトラック情報を扱えるようにした。

## 2. 1. 3. 2. ショウジョウバエ AGO2 結合低分子 RNA のハイスループットシーケンス解析

機能性 RNA データベースを活用した機能解析グループの慶應義塾大学塩見研究室との共同研究によって、ショウジョウバエの新規内在性 siRNA が発見された (Kawamura et al., 2008 *Nature* 453:793-797)。

塩見研究室では、マイクロ RNA や siRNA の作用経路に関与するタンパク質のひとつである AGO2 タンパク質に特異的に結合する抗体の合成に成功した。この抗体を用いてショウジョウバエの生殖細胞から、AGO2 タンパク質を免疫沈降法により精製し、このタンパク質に結合していた低分子 RNA をハイスループットシーケンシングによってシーケンス解析を行ったところ、77,327 本の低分子 RNA 配列情報を得た。バイオインフォマティクスグループでは、共同研究の一環として、この大量配列情報の情報解析を担当した。我々は機能性 RNA データベースを活用して、これらの大量配列情報の由来を自動的に判定する独自のアノテーションシステムを構築し (図 4)、迅速に配列の由来を決定することに成功した。

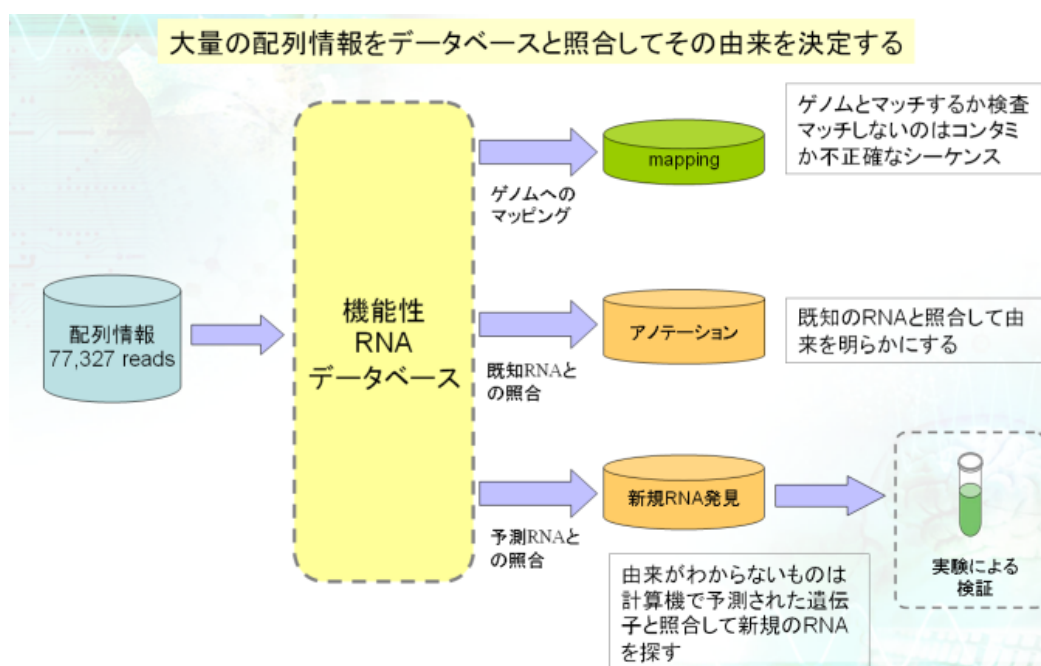


図 4. ショウジョウバエ AGO2 結合低分子 RNA の機能性 RNA データベースを用いた情報解析の流れ図

アノテーションシステムの処理の流れは次の通りである。配列情報は全てショウジョウバエゲノムへマッピングされる。完全一致の配列と 1 塩基のミスマッチを含むものだけを採用し、マッピングできなかった配列は後の解析には用いないようにして解析の信頼性を落とさないようにした。マッピングされた配列は、次にどの遺伝子や繰返し配列 (レトロトランスポゾンも含む) と重複しているか検査される。ここでの遺伝子とは、タンパク質遺伝子と機能性 RNA 遺伝子の両方を指す。まずは既知の遺伝子や繰返し配列との重複を検査し、次に予測された機能性 RNA 遺伝子との重複も検査する。図 5 に既知の遺伝子と判定

された低分子 RNA の内訳を示す。

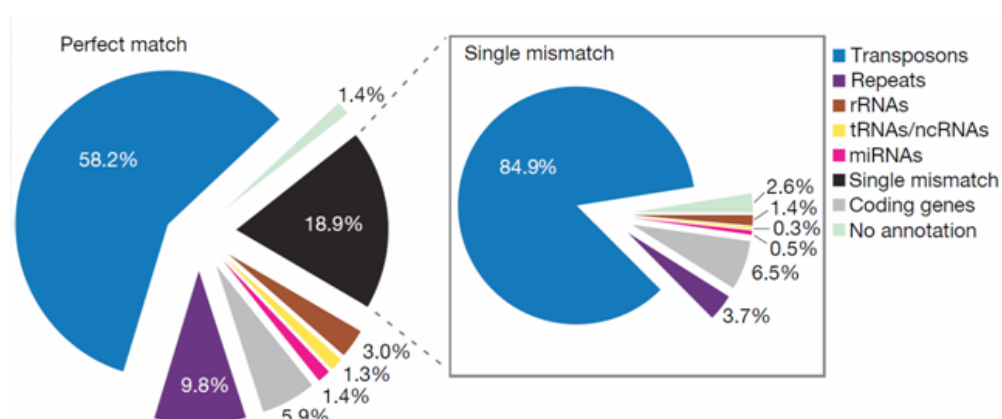


図 5. ショウジョウバエ AGO2 結合低分子 RNA のアノテーション結果

予測された機能性 RNA 遺伝子との重複によって、新規の機能性 RNA が自動的に検出可能である。いずれの遺伝子とも重複がなかったものについては、特に着目すべき特徴を持つものから順に手作業による検査を行い、2 個の新規マイクロ RNA を 2 つ発見した (図 6)。

どの遺伝子とも重複のなかった低分子 RNA のなかに、最も多く観測された低分子 RNA 配列が含まれていた。我々は、機能性 RNA データベースの一角である機能性 RNA ゲノムブラウザ UCSC GenomeBrowser for Functional RNA 上で、件の低分子 RNA 配列の由来を調べたところ、これらがゲノムの特定領域から発現していることがわかった。当該領域には転写産物がヘアピン構造をとるような配列が 20 回連続する特長的な領域であることを、機能性 RNA ゲノムブラウザ特有の機能を用いて瞬時に調べる事ができた (図 7)。このヘアピン構造が内在性 siRNA の母体となっていることが強く示唆された。他の競合グループも同様の発見を行ったが、我々は機能性 RNA データベースを用いる事によって迅速に作業を行ったことでより早く論文を投稿することに成功した。この実績によって、機能性 RNA データベースが低分子 RNA のハイスループットシーケンス解析において有用であることが明確に示されることとなった。

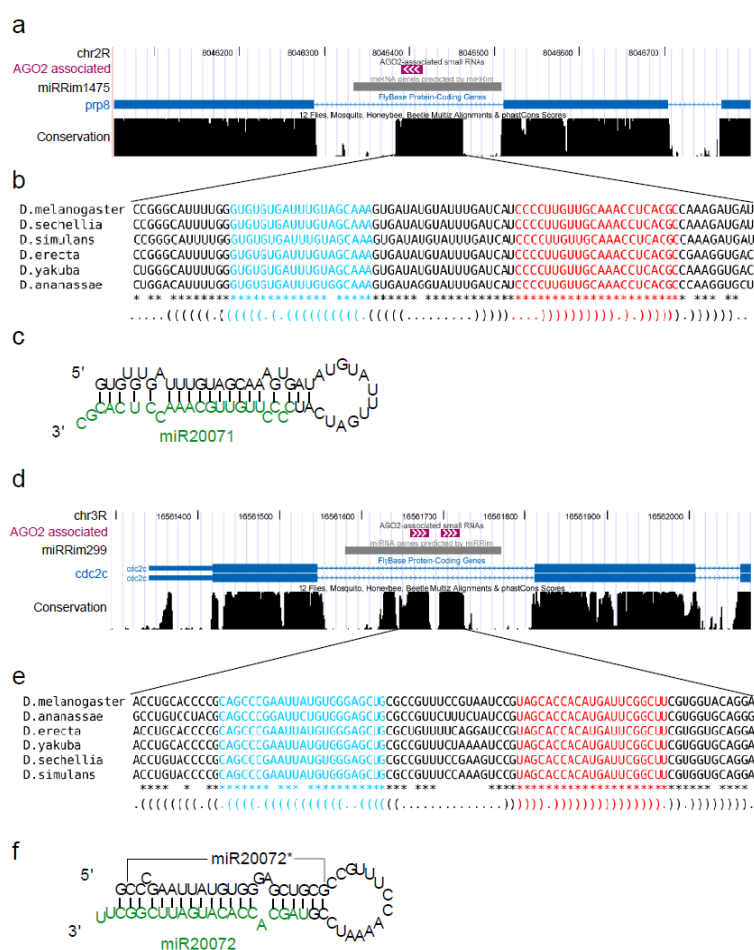


図 6. 機能性 RNA データベースを用いて発見された新規マイクロ RNA。一つは、我々独自のマイクロ RNA 予測手法である miRRim の予測によって得られたものである。

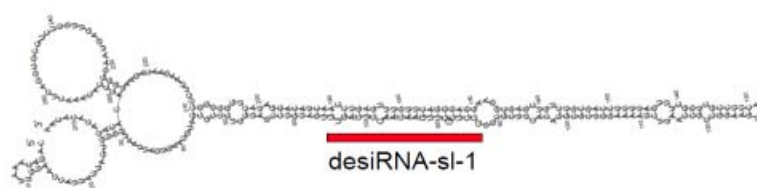


図 7. ショウジョウバエ AGO2 結合低分子 RNA の配列解析から発見された内在性 siRNA (赤線部) とその母体であるヘアピン構造の構造図

### 2. 1. 3. 3. ヒト AGO2 結合低分子 RNA のハイスループットシーケンシング解析

ヒト AGO2 タンパク質は、マイクロ RNA の生合成経路において、成熟マイクロ RNA を捕獲する重要な性質を持った RNA 結合タンパク質である。マイクロ RNA 合成機構の詳細を明らかにするためには、このタンパク質がどのような RNA を捕獲しているか詳細に調べる必要がある。

慶應義塾大学医学部塩見研究室（機能解析 Gr）実施のヒト AGO2 を標的にした IP-seq によって得られた配列情報からマイクロ RNA に相当する配列情報を選別する。マイクロ RNA 配列の性質を情報学的に解析することで、配列上の特徴を詳細に調査した。

その結果、マイクロ RNA には、最も頻繁に観測される regular form のほかに、異なる長さのものや塩基が付加されたバリエーションが多数観測された。特に 3' 端は 5' 端に比べて変化が大きかった（図 8）。さらに、新規マイクロ RNA を 5 件発見することができた（表 4）。そのうち 1 件は、①-2 で開発したマイクロ RNA 遺伝子予測プログラム miRRim で予測されていたものであった（これらは、マイクロ RNA の代表的データベースである miRBase に登録された）。上記の結果は、Azuma-Mukai A *et al.* 2008 *Proc Natl Acad Sci U S A* 105(23):7964-7969. に発表した。

表 4. miRBase に登録されている本プロジェクト成果による新規マイクロ RNA の一覧。hsa-mir-1538（赤丸）は、独自のマイクロ RNA 予測プログラムの miRRim によって検出できた新規マイクロ RNA

Accession	ID <sup>△</sup>	Literature reference	Links
<a href="#">MI0007261</a>	<a href="#">hsa-mir-103-1-as</a>	✓	PubMed ID: <a href="#">18524951</a>
<a href="#">MI0007262</a>	<a href="#">hsa-mir-103-2-as</a>	✓	PubMed ID: <a href="#">18524951</a>
<a href="#">MI0007258</a>	<a href="#">hsa-mir-1537</a>	✓	PubMed ID: <a href="#">18524951</a>
<a href="#">MI0007259</a>	<a href="#">hsa-mir-1538</a> ●	✓	PubMed ID: <a href="#">18524951</a>
<a href="#">MI0007260</a>	<a href="#">hsa-mir-1539</a>	✓	PubMed ID: <a href="#">18524951</a>

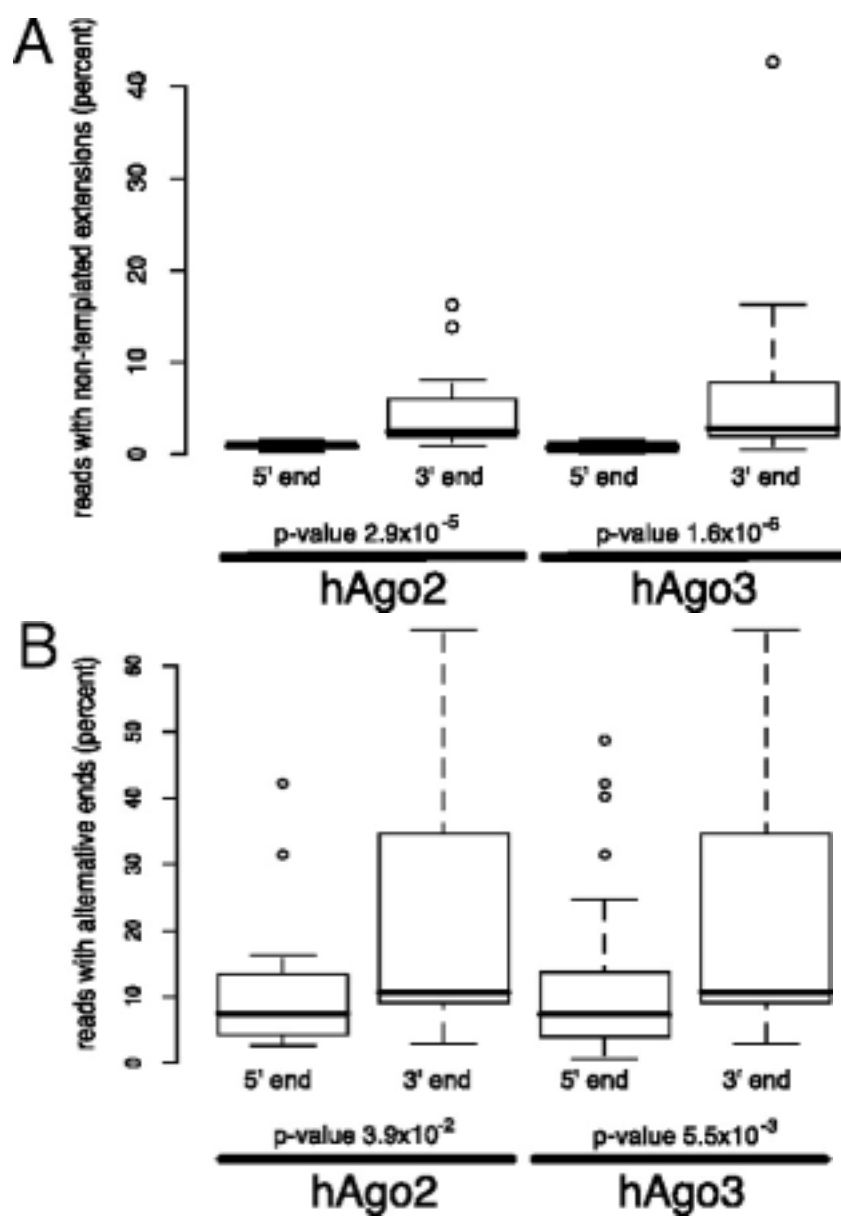


図8. A)マイクロ RNA の 5' 端及び 3' 端に、ゲノムに対応しない塩基が観測された頻度分布。B)Regular form に対する 5' 端、3' 端の長さの違いの分布。

## 2. 1. 3. 4. 機能性 RNA データベースの改訂

プロジェクトの中盤にあって、当初の絞込みによる候補は機能解析の段階に至ったが、新規機能性 RNA の探索と機能解析は、次世代シーケンサーの急速な普及によって大きく影響されることになった。完全長 cDNA シーケンシングのように一つ一つの転写産物をクローニングしてライブラリ化する段階を踏まずに、玉石混交の塩基配列を高速にシーケンシングして、情報解析によって詳細な内容を把握する手法が主流となった。このような状況が、機能性 RNA データベースを大量配列情報解析の情報基盤として発展させる契機となった。一方で、本プロジェクトを含め、世界全体での機能性 RNA 研究の進展に伴って、機能性 RNA の発見や機能解析に有用な情報が数多く発表されるようになった。これらの情報源に迅速に対応し、新規 RNA の発見と機能解析を支援するための、機能性 RNA データベースの改良及び拡張を行った。

機能性 RNA データベースは、配列データベースである「fRNAdb」とゲノムブラウザーである「UCSC Genome Browser for Functional RNA」が連携して動作する。

### 1) 配列データベース「fRNAdb」

既知の機能性 RNA 配列情報について、最新の情報に更新・追加する。登録配列は、次世代シーケンサーによる大規模シーケンス解析で得られた配列も含めて 510,055 配列となった。内訳は表 5 に示す。





表 5. 機能性 RNA データベース・バージョン 3 登録配列の内訳

種別	本数	内訳
既知 RNA	322,856	RNAdb Rfam miRBase NONCODE snoRNA-LBME-db European ribosomal RNA
予測 RNA	85,079	RNAdb (RNAz, EvoFold) H-invDB (NPCT)
大規模シーケンス	102,130	Gene Expression Omnibus
合計	510,055	

それぞれの登録配列について、配列の分類（オントロジー）に関する情報、塩基配列の特徴に関する指標、既知遺伝子との関連性、転写活性に関する情報を対応付けて、詳細画面にてそれらの情報が一覧できるようにした（図 9、10）。表示される属性の詳細について、表 6 に示す。また、大規模シーケンス解析によって得られた配列については、他の実

験との比較情報も表示する（同じ配列が他の実験で得られている場合）（図 1 1）。

fRNAdb の画面インタフェースを新しくし、表 6 に示したような詳細な属性について検索条件を絞り込むことが可能ないように、検索機能を強化した。図 1 2 に fRNAdb の画面を示す。fRNAdb のトップページ中段には 4 つのアイコンがあり、ここからデータベースの基本機能にアクセスできる。以下にそれぞれのアイコンの機能の詳細について示す：

	Catalog	登録配列の様々な属性（長さ、オントロジー分類、生物種など）による分類結果を示すことでデータベースの概要をつかみやすくし、あらゆる登録配列へのアクセスを容易にするためのインタフェース。
	Blast	NCBI Blast による配列相同性検索サービス
	Download	データベース情報のダウンロード画面
	Help	詳細なヘルプ情報（英文）

検索機能については、配列に関連付いた 26 項目について詳細な検索が可能ないように、それぞれの項目に対応したクオリファイア（修飾語句）が利用可能である。例えば、既知遺伝子の 3' UTR に重なるアンチセンス転写産物を検索するには、以下のように検索語を入力する：

%[ov\_anti] AND %[ov\_3utr]



表 6. 登録配列情報の詳細

分類	詳細項目	説明
基本情報	アクセション番号 オントロジー分類 由来生物種 ゲノムへのマッピング位置 外部 DB への相互参照 関連文献情報	
塩基配列の特徴	塩基配列 二次構造情報	既知構造が無ければ予測構造を表示する。予測構造では、進化的に保存された二次構造があればそれを表示し、そうでなければ単独の配列から予測した構造を表示する。
既知遺伝子との関連性	重複遺伝子との関係  隣接遺伝子との関係  OMIM 情報との関連	重複遺伝子とのセンス/アンチセンスの区別。重複領域が 5' UTR、エキソン、3' UTR、イントロンのいずれかの区別。 隣接（同じ転写クラスター※）内で、上流、下流のいずれかの区別。 ※転写クラスター=Affymetrix社の transcript cluster の定義に基づく。
転写活性に関する情報	Affymetrix GeneChip Invitrogen Ncode	配列上に各社のプラットフォームによるプローブが設計されていれば、その情報を表示する。

## FR251213

[Return]

Summary		Sequence	Secondary Structure
<b>Summary</b>			
ID	FR251213		
Description			
Accession	AL137603		
Sequence Ontology	antisense_RNA		
<b>Organism</b>			
Homo sapiens	human, man		
<b>Genome Mapping</b>			
human(hg18)	1 region	chrY:2909725-2910547(+) <a href="#">Browse</a> <a href="#">f</a>	
<b>Cross Reference</b>			
RNAdb v2.0	AS03699		
<b>Gene Association / Sense Overlap / 3'UTR</b>			
human(hg18)	ZFY <a href="#">uc004fqj.1</a> <a href="#">Browse</a> , <a href="#">uc004fqj.1</a> <a href="#">Browse</a> , <a href="#">uc010rwe.1</a> <a href="#">Browse</a> , <a href="#">uc010rwe.1</a> <a href="#">Browse</a>		
<b>Sequence Similarity</b>			
<b>OMIM</b>			
ID	Title	Locus	
490000 <a href="#">Browse</a>	ZINC FINGER PROTEIN, Y-LINKED; ZFY	Yp11.3(chrY:1-330000)	
<b>MicroArray / Affymetrix GeneChip Exon Array / hg18</b>			
Probe	Transcript Cluster	Probe Locus	
1760698	4028568 <a href="#">Browse</a>	chrY:2909858-2909862(+) <a href="#">Browse</a>	
5423723	4028568 <a href="#">Browse</a>	chrY:2909888-2909912(+) <a href="#">Browse</a>	
4697249	4028568 <a href="#">Browse</a>	chrY:2910400-2910424(+) <a href="#">Browse</a>	
<b>MicroArray / Invitrogen Ncode Noncoding RNA Array / hg18</b>			
Probe	Target Transcript	Probe Locus	
INGNh13844	AL137603 <a href="#">Browse</a>	chrY:2910019-2910079(+) <a href="#">Browse</a>	
<b>Reference</b>			
<p>1. RNAdb--a comprehensive mammalian noncoding RNA database.  <a href="#">Pang KC</a>, <a href="#">Stephen S</a>, <a href="#">Engström PG</a>, <a href="#">Tajul-Arfin K</a>, <a href="#">Chen W</a>, <a href="#">Wahlestedt C</a>, <a href="#">Lenhard B</a>, <a href="#">Hayashizaki Y</a>, <a href="#">Mattick JS</a>  <a href="#">Nucleic Acids Res</a>, 33(Database issue):D125-30, 2005 Jan  <a href="#">cited in PMC: 27</a></p>			

図9. 配列の詳細情報表示画面



## FR040099

[Return]

Experiment	Method	Tissue	Read Number	PubMed
small RNAs from 14 libraries in 2008/06/02 Organism: <i>Drosophila melanogaster</i>	Illumina-Solexa	Imaginal disc	271	1 citation
endogenous small RNAs bind to Argonaute2 in somatic cells in 2008/05/08 Organism: <i>Drosophila melanogaster</i>	Roche-454	S2 cells, AGO2 IP	1	1 citation
small RNAs from cell lines and imaginal disc in 2008/05/01 Organism: <i>Drosophila melanogaster</i>	Illumina-Solexa	Imaginal disc	271	1 citation
small RNAs in 2008/04/30	Illumina-Solexa	Wild Type adult female ovaries, 18-29nt	204	1 citation
	Illumina-Solexa	dcr-2 <sup>-/-</sup> mutant adult female ovaries, 18-29nt	324	1 citation
	Illumina-Solexa	loqs <sup>-/-</sup> mutant adult female ovaries, 18-29nt	421	1 citation

図 1 1. 大規模シーケンスによって得られた RNA 配列の発現情報を示す画面。実験情報、実験手法、組織、発現量（リード数）、参考文献が確認できるようになっている。

[Top](#)
[Catalog](#)
[Blast](#)
[Download](#)
[ncrna.org](#)
[Help](#)

## fRNAdb functional

A comprehensive non-coding RNA sequence database ver. 3.4

fRNAdb is Web Service (SOAP, REST) Ready.

Total: 510,055 entries

  
Catalog

  
Blast

  
Download

  
Help

Please input some query keywords for retrieving RNAs via simple text search e.g. "miRNA" or "snoRNA".  
Try add disease/tissue name to make your search more specific e.g. "miRNA oncogene".  
Or can be more specific e.g. "miRNA oncogene human".

[submit](#)

The icons used in this page are a part of Tango Desktop Project created by Tango Desktop Project which are available under CC BY-SA License.  
 This site uses Yahoo! User Interface Library (YUI) which is available under BSD License.  
 fRNAdb is licensed under a Creative Commons 表示-非営利-改変禁止 2.1 日本 License.  
 Based on a work at www.ncrna.org.

図 1 2. 配列データベース「fRNAdb」トップページ 中段の4つのアイコンからデータベースの基本機能にアクセスできる。「Catalog」は、登録配列のカatalog画面で、様々な属性による配列の分類

## 2) ゲノムブラウザ「UCSC GenomeBrowser for Functional RNA」

ゲノムブラウザでは、ゲノムに関連する様々な因子の位置情報を画面上に表示する。位置情報のまとまりを「トラック」とよぶが、ゲノムブラウザには多数のトラックが掲載されている。本プロジェクトのための追加したトラックの一覧を表7～9に示す。これら追加トラックの一覧は、ウェブ上でも公開しており、www.ncrna.orgのトップページから「Project Specific Custom Tracks」をクリックして閲覧可能になっている。

トラック追加に加えて、次に示す独自の機能拡張も行った。主な拡張は、動的な二次構造予測機能と、秘匿トラック情報機能である。

動的な二次構造予測機能は、ゲノムの任意の領域で、進化的に保存された二次構造を予測する機能である。比較ゲノムのトラック（Conservationトラック）をクリックすると、画面に表示されている領域の二次構造について予測し、その結果を表示することができる（図14, 15）。比較ゲノムのトラックをクリックするようにしたのは、二次構造予測において比較ゲノムの結果を考慮した二次構造予測を行っているためである。予測にはバイオインフォマティクスグループで独自に開発したCentroidFoldプログラムを用いている。結果画面には、比較ゲノムと塩基対の関係を対比しながら確認するための表示と、二次構造を表示するのに良く用いられるグラフ図を示す。比較ゲノムと塩基対の表示においては、ゲノム配列の置換と塩基対の関係を次のように色分けして表示している：灰色＝塩基対に関与しない塩基で種間置換を示さないもの、紫＝塩基対に関与しない塩基で置換を示すもの、黒＝種間で保存されている塩基対、青＝塩基対は保存されているが、片方の塩基に置換があるもの、緑＝塩基対は保存されているが、両方の塩基が置換したもの、赤＝塩基対を形成しなくなる単塩基の置換、橙＝塩基対を形成しなくなる2塩基の置換、マゼンタ＝塩基対を形成しなくなるような欠損を含む置換（図13）。

%ID	offset	0	1	2
	Human	GTTATTCAGGCCAAAAATGCCTGAATTAC		
86.7	Megabat	GTCATTCAGGCCAAAAATCACTGAATTAC		
86.5	Tenrec	GTCATTCAGGCCAAAAATCCTTGAATTAC		
85.5	GuineaPig	ATTATTCAGGCCAAACATGCTGAATTAC		
85.1	Sloth	GTTATTCAGGCTGAAAATGACT--ATTAC		
	SS anno	.....((((((.....)))))).....		
	pair symbol	abcde		edcba

図13. 比較ゲノムアラインメントと塩基対との関係を示す図。置換の種類によって色分けして表示する。

秘匿トラック情報機能は、ゲノムブラウザにユーザー認証機能を組み込むことによって、ユーザーごとにアクセスできる情報を制御することが可能になる。主には、共同研究に際して、共有したいが公開したくない情報をゲノムブラウザ上で取り扱うのに適している。

ゲノムブラウザには、カスタムトラックというユーザー独自のトラックを追加する機能があるが、これでは大規模な情報の取り扱いが難しい。また、公開サーバーを使わずに、非公開のサーバーに情報を格納し、VPNなどセキュアなネットワーク技術を使って情報共有する方法では、共同研究の数だけサーバーを立ち上げなければならない、その維持・管理の手間を考えると、複数の共同研究を同時に進行させるのは非現実的である。

表7. RNA 配列のマッピング結果と予測 RNA 情報のトラック (ヒト)

種別	内容	個数
分類済 RNA	tRNA, snoRNA, rRNA, snRNA, その他	6,652
予測 RNA	RNAz, EvoFold, snoSeeker, QRNA, intronic ncRNA, アンチセンス RNA, 機能解析グループの候補	270,800

表8. マイクロ RNA 関連のトラック (ヒト)

種別	内容	個数
既知 miRNA	miRBase	1,244
既知標的遺伝子	TarBase	233
予測 miRNA	miRRim, RNAmicro, Li2006, miRNAMap, Berezikov	11,616
予測標的部位	RNAhybrid, PicTar, miRBase Target, T-ScanS, PITA	2,384,525
miRNA 発現情報	Mammalian miRNA Expression Atlas	992

表9. 様々なゲノム因子のトラック (ヒト)

種別	内容	個数
進化的保存域	indel-based conservation, ultra-conserved, transposon-free	604,829
ヒストン修飾	H3K4me3	57,655
エンハンサー	Enhancer candidates, Tissue-specific enhancer candidates	7,748
cis 制御因子予測	Putative transcriptional cis-regulatory modules (PReMod)	123,510
核膜結合ドメイン	Lamina associated domains	1,344
スプライス部位予測	Predicted splice sites	1,058,924



### RNA Secondary Structure Prediction

● This view presents a result of evolutionary conserved RNA secondary structure predicted with **RNAalifold** from the corresponding multiple alignment.  
 The result is computed and presented automatically on any region shorter than 500 bp whenever the corresponding region has potential RNA secondary structure. However, this does not mean that there is certainly conserved RNA secondary structure. This feature is aiming to provide supportive information for functional RNA analysis.

#### Highlighting compensatory mutations in RNA multiple alignments

- Gray: Not part of annotated pair, no substitution.
- Light purple: Not part of annotated pair, substitution.
- Black: Compatible with annotated pair, no substitutions.
- Blue: Compatible with annotated pair, single substitution.
- Green: Compatible with annotated pair, double substitution.
- Red: Not compatible with annotated pair.
- Orange: Not compatible with annotated pair, double substitution.
- Magenta: Not compatible with annotated pair, involves gap.

strand	minimum free energy (kcal/mol)	# of pairs, no substitutions
forward	-89.54	68
reverse	-94.84	77

#### Multiple alignment and RNA 2ndary structure (forward strand)

```

>ID offset 0 1 2 3 4 5 6
Human CTGGGAAAGGSGCCAAAGCCAGGGTSAATCACTTTCTTAAAAAATAAGCTCAAGTGGAGTCCAGAGG
97.4 Chimp CTGGGAAAGGSGCCAAAGCCAGGGTSAATCACTTTCTTAAAAAATAAGCTCAAGTGGAGTCCAGAGG
96.9 Gorilla CTGGGAAAGGSGCCAAAGCCAGGGTSAATCACTTTCTTAAAAAATAAGCTCAAGTGGAGTCCAGAGG
SE anno .-(((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
pair symbol abcdefghijklm nopqrst uvwxyz ABCDE EDCR DCBFG

offset 7 8 9 0 1 2 3
Human GAATGGGASAAAACCCTCCCGCTCCCTGTGTGGAAGTGTAGGAGCCGCCACCCCHAACTACTCHACTH
Chimp GAATGGGASAAAACCCTCCCGCTCCCTGTGTGGAAGTGTAGGAGCCGCCACCCCHAACTACTC--ACC
Gorilla GAATGGGASAAAACCCTCCCGCTCCCTGTGTGGAAGTGTAGGAGCCGCCACCCCHAACTACTC--ACC
SE anno .-((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
pair symbol H IJKLMNO ONHLSKLM NOPQ R R QPQHEKEL ELHM

offset 4 5 6 7 8 9 0
Human TGAATCCAGCCCTCTGGGAAAGGAGGGGTCATGAACTCCGCCCTAGTCACAGGGGCTCCCTGTGGC
Chimp TGAATCCAGCCCTCTGGGAAAGGAGGGGTCATGAACTCCGCCCTAGTCACAGGGGCTCCCTGTGGC
Gorilla TGAATCCAGCCCTCTGGGAAAGGAGGGGTCATGAACTCCGCCCTAGTCACAGGGGCTCCCTGTGGC
SE anno .-((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
pair symbol MNR OI HGFEDCRkqvwxyz uvwxyz yzvwxyzq rstuvw vutxyzqo

offset 1 2 3
Human CCAAGGCCCTCTTCACTCCAT
Chimp CCAAGGCCCTCTTCACTCCAC
Gorilla CCAAGGCCCTCTTCACTCCAT
SE anno 1....))))))))....)))))
pair symbol n mkljshgfe dcba
    
```

view the alignment in [Stockholm format](#) or [ClustalW format](#).

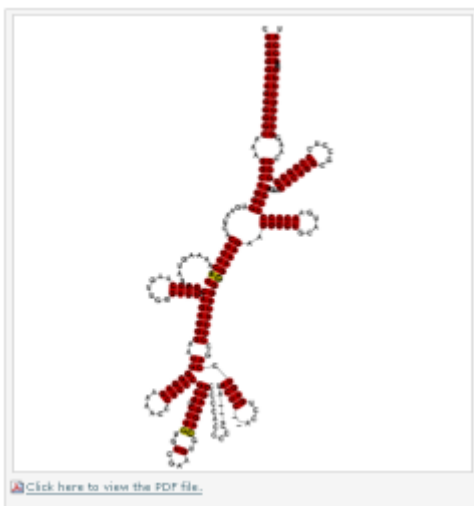


図 1 5 . 動的二次構造予測結果の表示画面 + / - の両側のストランドにおける二次構造を予測する。各ストランドで予測された塩基対の数と最小自由エネルギー (MFE) が示されるので、どちらのストランドでより安定な二次構造が形成されるかを判断する材料を提供する。



### 2. 1. 3. 5. ヒト RNA 編集領域の網羅的解析

東京大学鈴木研究室（ツール開発 Gr）との共同研究により、RNA 転写産物から、アデニンがイノシン酸になる RNA 編集（イノシン酸はシークエンサーによってグアニンとして観測される）が観測される部位をゲノムワイドに計測し、RNA 編集の機能についてマイクロ RNA との相互作用の側面から情報学的な検証を行った。

鈴木研究室では、ヒトの培養細胞を用いて、RNA 転写産物から RNA 編集が生じている部位を網羅的に測定した。測定結果を機能性 RNA データベースに登録し、RNA 編集がマイクロ RNA の標的領域としてどのような影響を与えるかを情報学的に検定するため、マイクロ RNA 標的予測プログラムである Miranda を用いて、RNA 編集部位が関与するマイクロ RNA 標的部位を網羅的に列挙した。エディティングがある場合とない場合とで、マイクロ RNA 結合エネルギーの変化量を網羅的に予測した。予測に際しては、イノシン酸をグアニンとして扱った。後にイノシン酸をグアニンとして扱うと、結合エネルギーの計算において不正確となる問題が判明したため、現在イノシン酸に対応するエネルギーパラメータを用いて結合エネルギーの再評価を行っている最中である。

測定情報を東大本郷キャンパスから機能性 RNA データベースにオンラインでアップロードすると直ちに機能性 RNA ゲノムブラウザで閲覧できるような機能を開発した（図 1 6）。アップロードと閲覧にはあらかじめ登録したユーザー名とパスワードが必要となっており、セキュリティにも配慮した設計になっている。

マイクロ RNA の標的部位に対する影響の評価は現在も継続中である。



図 1 6. RNA 編集情報のアップロード画面

### 2. 1. 3. 6. 自動アノテーションシステムの開発

次世代シークエンサーによる大規模な転写産物（RNA）の解析に対応するため、機能性 RNA データベースを応用した、自動アノテーションシステムの開発を行った（未発表）。これは、次世代シークエンサーから出力された配列情報をインターネット経由でアップロードすると、人手を介在させずに自動的にゲノムへマッピングし、アノテーションを付与するシステムである。マッピングやアノテーションの処理は、4-2. や 4-3. で実施した手順

を踏襲・改良したものである。

プロジェクト終了までにプロジェクト内で公開することを目指したが、平成22年2月末の時点でグループ内における試験運用段階にある。

図17に、自動アノテーションシステムの現時点のウェブ・インタフェース画面を示す。

図17. 自動アノテーションシステムの画面。「Sequence file」にシークエンサーからの配列ファイル名を指定し、ファイルの形式（fasta, fastq 形式に対応）を選択し、生物種を指定し、「Next」をクリックすると、ファイルがアップロードされ、アノテーションのバッチ処理が実行される。

### 2. 1. 3. 7. まとめ

機能性RNAデータベースの構築では、新規機能性RNAの探索と解析を支える強力な情報基盤となるデータベースを構築することを課題として実施した。プロジェクト完了後にデータベースが使えるようになるのでは無く、プロジェクト期間内に迅速にデータベースを構築し、有効に活用されることを意識した。さらに、プロジェクト完了後には、本プロジェクトの成果を発信する役割に加え、データベースが広く産業利用されうるような幅広い要求に応えられるものを提供するよう意識したことから、初期の研究に対応したバージョン1、中期～後期の研究に対応したバージョン3、プロジェクト終了後の産業利用を意識した「自動アノテーションシステム」と、大きく3つの開発フェーズを経ることとなった。我々は、単にデータベースの構築のみに終始するのではなく、データベースを利用した様々な情報解析も実施することで、実際の研究に役立つ情報ツールとしてのデータベースの構築を実践した。それによって、グループ間の連携を促進し、プロジェクト全体研究推進に直接的に貢献することができたと自負している。

## 2. 2 研究開発項目② 機能性RNA解析のための支援技術・ツールの開発

### 【成果の概要】

機能性 RNA 研究を強力に推進していくためには、新しい技術論・方法論の導入が求められている。本研究開発項目では、機能性 RNA を迅速かつ定量的に、そして網羅的に解析できる支援技術・ツール開発を行う。特に、発生や分化、疾患により変動する RNA を解析するため、測定感度を極限まで高めることで、細胞から抽出した微量な機能性 RNA を直接、測定する技術・手法を開発する。中間目標として、機能性 RNA を網羅的にかつ定量的に解析できるバイオツールの開発を目指し、平成 21 年度の最終目標としてサブフェムトモルオーダーでの直接解析を掲げた。

RNA マススペクトロメトリー (RNA-MS) の開発では、生体から抽出した微量な機能性 RNA を高感度質量分析法によって直接解析することを目指している。転写後プロセシングや修飾など、RNA が有する質的な情報を正確に読み取ることによって、RNA が関与する高次生命現象の解明や疾患との関連性を明らかにしていくことを主眼としている。プロジェクト開始 2 年目には、RNA-MS による高感度測定系の構築に成功し、最終目標値であったサブフェムトモルオーダーでの直接解析をクリアした。この成果が高く評価され、平成 19 年度には加速予算を投入いただき、本開発テーマの進展を加速させることができた。また、RNA-MS を支えるための周辺ツールとして、RNA-MS によって得られる膨大な質量情報を情報処理するためのアルゴリズムやソフトウェアの開発にも成功し、知財戦略の面でも優位に立った。さらに、世界初の全自動 RNA 精製装置 (往復循環クロマトグラフィー) の開発に成功し、細胞内に存在する微量な RNA の解析技術が飛躍的に向上した。RNA-MS は今や、機能性 RNA 研究のみならず、創薬や核酸医薬の開発を強力にサポートする基盤技術に成長した。これらの手法を駆使し、piRNA の末端修飾構造の同定や miRNA の直接プロファイリングなど、これまでの手法では不可能であった解析に成功している。

RNA 医薬品の開発を目的とした RNA の新規合成基盤技術開発では、高品質かつ安価な RNA の化学合成技術を確立することに成功した。実際に、この手法により、長鎖 RNA (110 塩基) の化学合成に成功し、生物活性を確認することができた。高品質かつ安価な合成 RNA を供給するシステムが整いつつあり、実用化と事業化に大きく前進した。

機能性 RNA の検出・同定技術の開発では、超高感度 (アトモルレベル) まで高めたマイクロアレイ技術を開発し、機能性 RNA の高精度な発現変動解析を目標としている。日本発でオリジナリティの高い多重伸長反応法および光クロスリンク法を駆使した RNA の高感度検出法を開発した。

**【研究開発方針】**

機能性 RNA 研究を強力に推進していくためには、新しい技術論・方法論の導入が求められている。本研究開発項目では、機能性 RNA を迅速かつ定量的に、そして網羅的に解析できる支援技術・ツール開発を行う。特に、発生や分化、疾患により変動する RNA を解析するため、測定感度を極限まで高めることで、細胞から抽出した微量な機能性 RNA を直接、測定する技術・手法を開発する。最終目標として、機能性 RNA を高感度(サブフェムトモルオーダー)で定量的かつ直接測定する手法の開発を掲げた。機能性 RNA の検出・同定技術の開発では、東京大学と島津製作所が担当する RNA マススペクトロメトリー (RNA-MS) は、生体から抽出した微量な機能性 RNA を高感度質量分析法によって直接解析する技術である。この手法では、微量な RNA を cDNA 化や標識をせずに生のまま解析することができる。細胞や組織から単離あるいは分画した微量な RNA をそのまま、あるいは断片化し、高感度質量分析法 (MS) により RNA 断片の質量を精確に測定することで、各断片に存在する RNA 修飾や末端構造の詳細な解析が可能となる。RNA が有する質的な情報を正確に読み取ることによって、RNA が関与する高次生命現象の解明や疾患との関連性を明らかにしていくことが期待される。また、miRNA を直接的に解析することによる診断技術の開発や、核酸医薬の薬物動態にも応用することが期待される。

オリンパス、北陸先端大は光クロスリンク法を RNA の検出に応用し、超高感度(アトモルレベル)まで高めたマイクロアレイ技術を開発する。さらに、武庫川女子大と DNA チップ研究所のグループは全く新しい発想に基づく多重伸長反応法 (MPEX 法) の開発に取り組み、実用的なマイクロ RNA のプロファイリング技術の確立を目標とする。

RNA 医薬品の開発において、日本が優位に立つために、高品質な RNA の化学合成技術が不可欠である。日本新薬と東京大学のグループは、革新的な核酸合成の技術を駆使し、実用的な RNA 合成法の確立と事業化を目標とする。

## 【各研究室における成果】

### 2. 2. 1 RNA マススペクトロメトリー法の開発

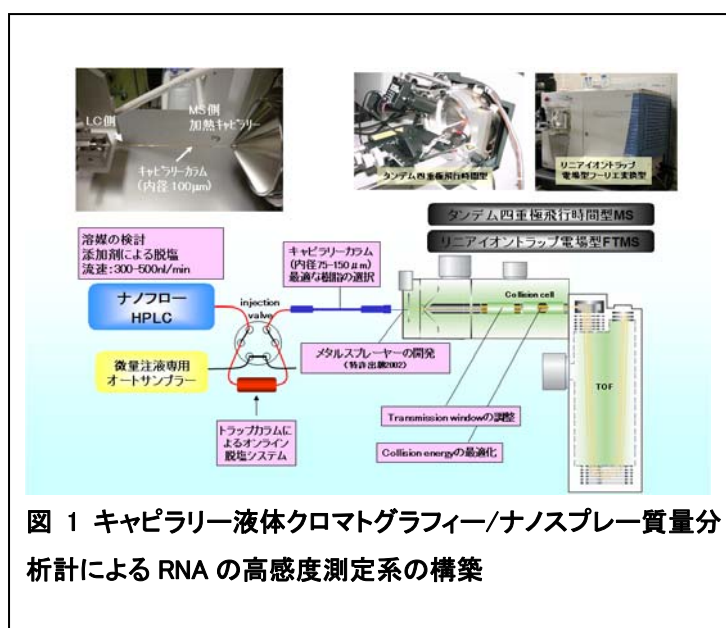
集中研②、分室4（島津製作所）

共同実施先： 東京大学（2；鈴木）

本研究開発項目では、RNA を“情報”として捉える従来型の解析方法から脱却し、RNA 分子を“もの”として捉える新しい方法論として、微量な機能性 RNA を直接的かつ定量的に解析する RNA マススペクトロメトリー（RNA-MS）の開発を行っている。通常、微量な RNA の解析法としては、逆転写と PCR を組み合わせた cDNA の解析が一般的であるが、ランダムプライミングやリンカーライゲーションの効率や PCR 増幅による cDNA のバイアスが生じるため、定量性に欠けるなどの問題点が指摘されている。さらに、末端構造や塩基修飾など、機能性 RNA が有する質的な情報を読み取ることができないという本質的な欠点がある。また、クローニングやシーケンスを含めると解析にかかる時間も考慮に入れなければならない。RNA-MS は、細胞及び組織より調製した微量な RNA を直接解析することができる。電気泳動や免疫沈降法によって単離あるいは分画した RNA を断片化し、高感度質量分析法によって解析することによって、断片の質量を正確に測定することができるため、各断片に存在する RNA 修飾の解析や、末端構造の詳細な解析が可能である。さらに、各 RNA 断片の分子量を正確に測定することで、シーケンス解析をすることなく、in silico 解析によって RNA 遺伝子の配列をゲノム上から迅速に同定すること（RNA マスフィンガープリント）も可能である。本基盤技術の開発は、RNA が関与する高次生命現象の解明や疾患との関連性を明らかにしていくことのみならず、将来的な RNA 創薬や再生医療、あるいは RNA を分子マーカーとした新しい診断技術へと応用が期待される。

#### 2. 2. 1. 1. キャピラリー液体クロマトグラフィー/ナノスプレー質量分析計による RNA-MS の構築

当初、我々はプロジェクトの最終目標値として、数フェムtomolオーダーでの RNA の微量解析を掲げた。プロジェクト開始前の測定感度が数ピコmolであったことから、測定感度を3桁向上させることを目指したものである。平成17年度には、タンデム四重極飛行時間型質量分析装置の測定パラメータの最適化、キャピラリーナノフロー液体クロマトグラフィーシ



システムの構築（図 1）、トラップカラムを用いたオンライン脱塩システムの構築、溶離液と分離樹脂の検討、イオン源の最適化などを徹底的に行い、その結果、内径 75-150 $\mu$ m のキャピラリーカラムを用いて、流速 300-500nL/min で送液を行うことで、1-5 フェモトモルの RNA を高感度で測定することに成功しプロジェクトの最終目標値を 1 年目にしてクリアするに至った。

平成 18 年度は、更なる高感度化を目指し、解析システムの微調整と条件設定を重点的にを行い、更にはキレート剤添加によるサンプルの前処理技術の検討を行い、50 アトモル(0.05 フェモトモル)の測定感度を達成した（図 2）。この成果は RNA を増幅や標識をすることなく、直接的に測定することに成功した世界記録であり、PCR やハイブリダイゼーションに頼らない全く新しい解析技術 RNA-MS が誕生したことを意味している。加速財源の獲得にも成功し、質量精度の高い電場式フーリエ変換型質量分析装置(Orbitrap)を導入し、miRNA の直接解析などに活路を見出した。

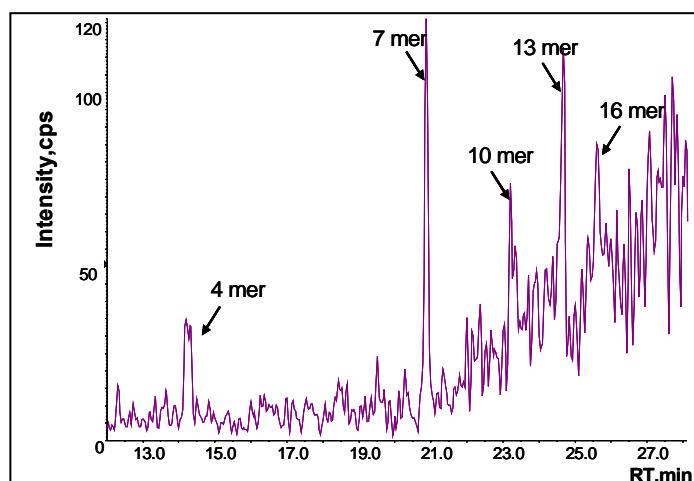


図 2 50 アトモルの世界最高感度を達成した

また RNA-MS 解析の膨大なデータを情報処理するためのツール開発を積極的に行った。RNA-MS デコンボリューション（著作物登録）（図 3）は、RNA の多価イオンを読み取り変換し、質量情報を算出するプログラムである。測定データを XML 形式に変換し、天然同位体分布を考慮した MS スペクトルのフィッティングと多価イオンをデコンボリューションすることにより、各 RNA 分子の質量を算出することを基本原理とする。また、RNA-MS

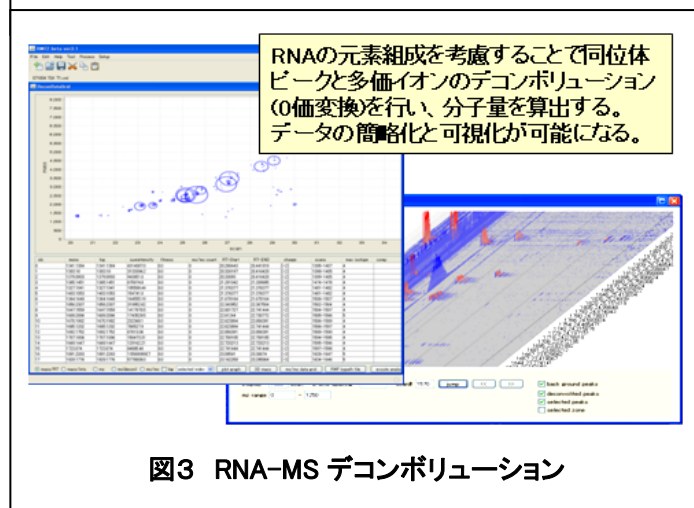


図3 RNA-MS デコンボリューション

による *de novo* シーケンスを強力にサポートする CID チェッカー（国際出願済）や、質量情報のみから計算科学的に RNA 遺伝子をゲノム上から同定する RNA マスフィンガープリント法（出願済）などを開発し、知的財産の戦略面でも優位に立っている。

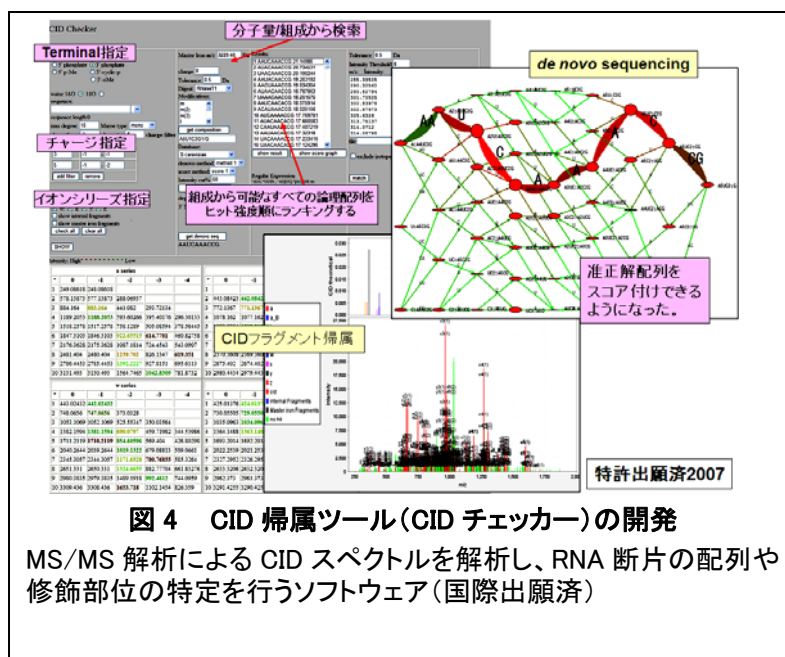
我々は RNA-MS を機能性 RNA の解析に積極的に活用している。機能未知遺伝子群から網羅的に RNA 修飾遺伝子を探索するアプローチ(リボヌクレオーム解析)や、RNA 修飾異常に起因する疾患の研究、RNA 修飾が決定する tRNA の細胞内局在化機構、ncRNA に含まれる RNA 修

飾の同定などは RNA マススペクトロメトリーの長所を生かした研究例である(業績参照)。特筆すべき成果として、マウス精巢に発現している Piwi- interacting RNA(piRNA)の3'末端が、ほぼ 100%修飾されていることを発見した(*Nat Struct Mol Biol.*, 2007)ことや、肝臓特異的に発現する miR-122 の 3' 末端がアデニル化されることで選択的に安定化される機構を見出した(*Genes Dev.*, 2009)ことなどが挙げられる。また、アーキアの tRNA から新規 RNA 修飾、アグマチジン(2-agmatinylcytidine)を発見し、生合成機構および機能解析に成功した(*Nat Chem Bio.*, 2010)。

## 2. 2. 1. 2. RNA-MS による配列解析技術と RNA マスフィンガープリント(RMF)法の開発

イオン化した RNA 分子は質量分析装置内で希ガスと衝突させることにより、解裂させることができる(CID=collision-induced dissociation)。CIDによって得られた内部配列に由来するプロダクトイオンを帰属することで、RNAの配列を de novo で決定することが可能である。通常、RNAの多価イオンから生じたプロダクトイオンは非常に複雑であり、人が目で見て解釈するのは困難を極める。そこで我々はCIDスペクトルを帰属するためのツールとして、CIDチェッカーを開発した(図4)。このアルゴリズムでは、塩基配列を探索するための有向グラフを作成し、有向グラフ上のノードとエッジを質量スペクトルに基づいて重み付けし、動的計画法を適応して、有向グラフ上の最適経路をたどることで塩基配列を解析している。さらに、RNA修飾を含んだ断片を解析することで修飾部位の特定にも威力を発揮する。

RNA マスフィンガープリント(RMF)法は RNA-MS で取得した質量情報を用い、RNA がコードされている遺伝子領域をゲノム上から in silico で探索するアルゴリズムである。RNA 断片の分子量特性を詳細に検討した結果、個々の RNA 断片の分子量はユニークな値を示すことが判明し、0.2 ダルトン以下の質量精度で測定すれば、個々の RNA 断片の塩基組成を一義的に決定できることが判明した。この知見を元に、我々は、RNA 断片の質量をクエリーとしてゲノム配列の表裏からその RNA をコードする領域を in silico で探索するアルゴリズム



(Genomic RMF)を開発した(図5)(特許出願済)。大腸菌、酵母に加えて、ヒトとマウスのゲノムから検索可能な Genomic RMF を開発し、実際にヒトとマウスの ncRNA の解析データを用いて、ゲノムの表裏(6 Gbp)上から遺伝子を特定することに成功している。また、CID チェッカーと組み合わせることにより、部分的に決定された配列情報を元に RNA 遺伝子の同定精度を飛躍的に向上させることが可能である。

RMF 法は、RNA の増幅や標識をすることなく、生の RNA を直接的に解析し、迅速に遺伝子領域を特定

することができる。この手法のアプリケーションとしては、RNA-タンパク質複合体(RNP)中に含まれる RNA 成分の同定が挙げられる。機能未知な RNA 結合タンパク質を免疫沈降法で精製し、結合している RNA を迅速に同定することで、新規の RNA-タンパク質の相互作用を明らかにすることができる。また、これまでによく調べられている RNP を解析することで、既知の RNA 以外に未知の RNA が見つかる可能性も大いに期待できる。実際に出芽酵母(*Saccharomyces cerevisiae*)の発現ライブラリーを用い、いくつかの RNP の解析を行い、実際に RNA 遺伝子が同定できることを確認した。また、産総研で整備されたヒトの完全長 cDNA を用いて、機能未知な RNA 結合タンパク質をヒト培養細胞で発現させ、結合した RNA 成分の同定にも成功している。MS 解析から Genomic RMF による遺伝子の特定にかかる時間はわずか1時間程度であり、この手法は網羅的な RNA-タンパク質の相互作用ネットワークの解析に威力を発揮することが期待される。RNP の精製には、後述する往復循環クロマトグラフィーを用いることにより、多検体 RNP をハイスループットに精製し、網羅的な解析システムを構築していきたいと考えている。

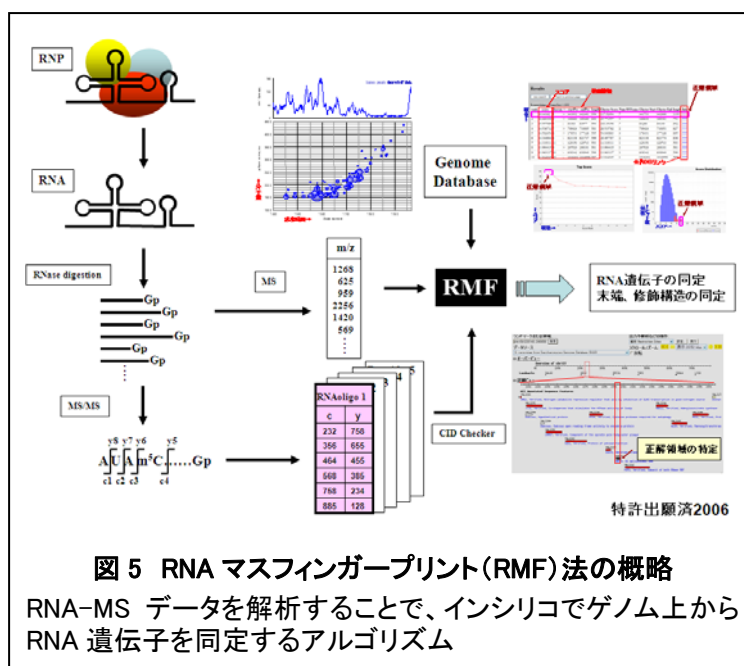


図5 RNA マスフィンガープリント(RMF)法の概略

RNA-MS データを解析することで、インシリコでゲノム上から RNA 遺伝子を同定するアルゴリズム

### 2. 2. 1. 3. マイクロ RNA の直接解析系の構築

miRNA の検出には、数多くの方法が開発されているが、定量リアルタイム PCR (qRT-PCR) 法やマイクロアレイを用いた手法が主流になりつつある。最近では、次世代シーケンサーによる低分子 RNA の大量解析がマイクロアレイに取って代わる手法になりつつある。これらの手法は miRNA の網羅的な発現プロファイルを解析する手段としては実用的ではあるものの、逆転写酵素による cDNA への変換や、PCR による増幅、蛍光標識による増感といった間接的な検出方法に頼っているのが現状であり、これらの過程で生じるバイアスは定量解析



を行う上で決して無視できない問題である。また、cDNA 化や標識などの作業にかかる労力や時間も考慮に入れなければならない。さらに、miRNA には多くのプロセシングバリエーションが存在することが知られており、バリエーションごとに標的 mRNA が変化する可能性があることから、miRNA をバリエーションごとに分離して検出する必要がある。しかし、従来法では 1 塩基の違いのバリエーションを区別して検出するのが難しいという問題もある。我々は、従来法では解決することが困難な問題点を克服するために、RNA-MS を用いて miRNA を直接解析する系の構築を目指した。

miRNA は 6000-8000Da の分子量分布を有し、nanoESI 法でイオン化すると、マスペクトル上では、通常、7~10 価程度の多価イオンとして検出される。また、この程度の大きさの分子量になると個々のマスペクトルが天然同位体（主に  $^{13}\text{C}$ ）に由来する複雑な同位体イオンによって占められており、これらの複雑なスペクトルを解読し、元の分子量を算出する必要がある。市販のプログラムでは RNA のスペクトルに対応していないため、独自に開発した RNA デコンボリューション（東大著作物）を用い、スペクトルから分子量を算出（0 価へ変換する）ことができるようになった。こうして得られた分子量は、通常は（修飾がなければ）、RNA の 4 つの塩基の組み合わせで計算することができるため、miRBase に登録されている miRNA の分子量と突合せを行うことで候補を割り出すことが可能である。更に、同じ塩基組成の miRNA を区別するためには、イオン化した RNA を CID によって分解し、生じたプロダクトイオンを CID チェッカー（国際出願）によって読み取ることによって、RNA の内部配列を決めることが可能である。

実際に、マウス臓器由来の miRNA について解析を行った。肝臓、脳、心臓から Total RNA を抽出し、定法にしたがって miRNA 画分を得た。約 3 pmol 分（肝臓の約 80mg に相当）の miRNA 画分を RNA-MS によって解析を行った。測定データは RNA デコンボリューションによりモノアイソトピック質量に変換した。また、いくつかの miRNA は CID 解析により部分配列を得た。各臓器から 100~200 個のピークが観測され、各分子量および CID によって得られた部分配列を、miRBase に登録されたマウス miRNA に対し参照した。各 miRNA にはプロセシングの位置が異なる両末端のバリエーションや、3' 末端に A や U が酵素的に付加されたバリエーションが存在するため、登録されている個々の miRNA ごとに末端バリエーションを考慮し検索を行った。その結果、RNA-MS で観測された miRNA 集団の中には、データベースに登録されていないバリエーションが多数含まれることが判明した。また、臓器ごとに特異的な miRNA が発現していることも確認された。また、分子量および CID のみから判定した miRNA の帰属が正しいかどうかを、検証するために、RNA-MS で解析に用いた miRNA 画分をディープシーケンスによって解析を行った。その結果、RNA-MS で帰属した全ての miRNA（バリエーションを含む）を確認したことから、本手法が信頼性の高い技術であることを示すことができた。

RNA-MS 解析によって観測された miRNA の各ピークの高さは、miRNA の存在量を反映しているが、マスペクトルのピーク強度は、各分子のイオン化効率や同時にイオン化される他の分子の影響を受けるため、ピーク強度から単純に存在量を算出することはできない。

一般的にマスマスペクトロメトリーによる定量解析では、 $^{15}\text{N}$  や  $^{13}\text{C}$  などの安定同位体で標識した標品を内部標準物質として定量する手法が用いられている。実際に、プロテオミクス研究においては、SILAC 法や iTRAQ 法などに代表される安定同位体標識を利用したタンパク質の発現定量が盛んに行われている。RNA-MS においても、安定同位体標識した RNA を用いることで各ピークの定量解析が可能である。実際に、マウス肝臓で高発現している miR-122 の 4 種類のバリエーションに対する安定同位体 RNA を合成し、適当量を内部標準として、肝臓由来 miRNA 画分に混合して測定を行った。安定同位体 RNA は天然の RNA とその化学構造が同一であり、キャピラリー LC から同時に溶出されてくるため、イオン化の効率が完全に同じである。マスマスペクトル上では分子量が異なるためにそのピーク強度比から、天然 RNA の絶対量を見積もることが可能である。実際に、miR-122 の 4 種類のバリエーションの定常状態量を高い精度で測定することに成功している。通常、RNA の発現定量は、定量的逆転写 PCR (qRT-PCR) 法などに代表されるような相対定量法が主流であるが、安定同位体 RNA を内部標準として用いる方法は、分析化学の分野で用いられる精度と信頼性の高い絶対定量法である。このような定量法は、特に疾患などで発現量や定常状態量が変動する miRNA の解析に威力を発揮すると期待している。

ここ数年、固形癌や白血病をはじめとする様々な疾患で miRNA の発現変動が報告されている。また、血漿中に存在するエクソソーム (exosome) に miRNA が内包されており、細胞がエクソソームを介して、転移や細胞増殖に関わる情報をやり取りしている可能性が指摘されている。エクソソーム中に含まれる miRNA を調べることで、診断のみならず癌細胞の種類なども特定できる可能性があると考えられている。miRNA は細胞分化の指標であり、miRNA を定量的にプロファイリングすることで、様々な疾患の診断マーカーとして活用できると期待される。我々は、現在 RNA-MS を用いてヒト血清に含まれる miRNA のプロファイリング技術の確立を目指している。

#### 2. 2. 1. 4. MALDI 型質量分析計による RNA の高感度測定系の構築

MALDI 型質量分析計は、操作が簡便で測定も迅速に行えることから、複雑な混合物の中から、特定の分子マーカーを検出したり、全体のプロファイリングを取得する測定系に適している。微量な RNA を MALDI 法でイオン化し、高感度で測定するためには、RNA を高効率でイオン化するためのマトリクスの開発が不可欠であった。従来オリゴ核酸はプロトンが付加された正イオンとして検出されてきたが、我々は、RNA が酸性分子であることの特徴を生かし、RNA を負イオンとして検出することを指標にしてマトリクスを探索した。その結果、3-Hydroxypicolinic acid (3-HPA) に添加剤として Ammonium citrate dibasic (DAC) を用いた系で最適なマトリクスを見出し、数フェムトモルオーダーの RNA をイオン化することに成功した (特許出願済、島津で製品化の予定)。特に、MALDI 法では出にくいことが知られている、 $-2$  価のイオンを検出することに成功し、より高質量側での測定に活路を見出すことができた。

この手法を用い、マウス臓器由来のマイクロ RNA 画分を測定したところ、分子量 7000Da 付近に個々のマイクロ RNA 由来のシグナルが多数観測された (図 6)。正確な分子量と CID 解析の結果、個々のシグナルを各 miRNA 分子に帰属することに成功した。臓器ごとに発現している miRNA のサブセットは異なるパターンを示し、マイクロアレイや cDNA クローニングで文献的に報告されているデータとよい一致を示した。検出限界を測定したところ、臓器 2mg 由来の miRNA 画分 (43 fmol) でいくつかのメジャーな miRNA を検出できた。また、検量線を用いることで、個々の miRNA を定量的に解析することができる。この方法が実用化されれば、生検サンプルや摘出した臓器由来の miRNA 画分を迅速にかつ定量的に測定することができるため具体的な診断技術として確立できる可能性が高い。

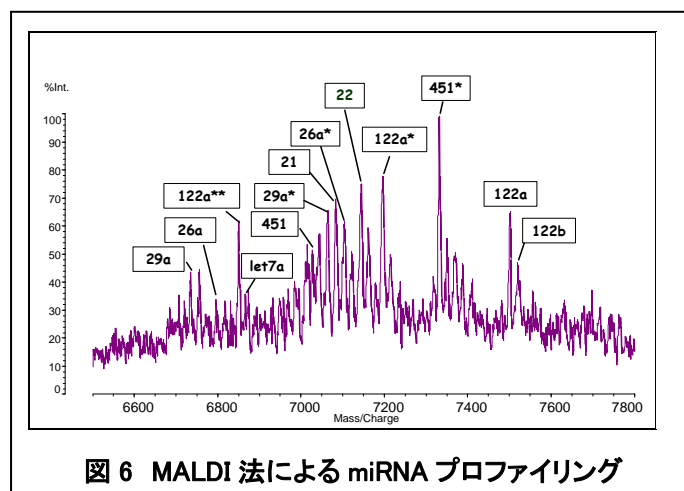


図 6 MALDI 法による miRNA プロファイリング

## 2. 2. 1. 5. 往復循環クロマトグラフィーを用いた全自動 RNA 精製装置の開発

微量な RNA 分子の単離精製は一般的に困難を極め、経験値やノウハウに左右される難易度の高い技術である。我々は長年 RNA 研究に携わってきた経験から、細胞内に存在する微量 RNA を単離精製する技術を開発してきた。往復循環クロマトグラフィー (*Nucleic Acids Res.*, 2007) (国際特許出願済) は全く新しい発想から生まれた多検体アフィニティークロマトグラフィーである (図 7)。マルチピペッターが搭載された自動分注機を用い、並列に並べたアフィニティーチップで同時に吸引、吐出、攪拌を繰り返すことで、全ての試料溶液を全てのアフィニティーチップに均一に循環させることを基本原理としている。異なる標的分子に対する複数のアフィニティーチップ (RNA の精製には DNA 固相化樹脂) をマルチピペッターに装着することで、試料溶液を同時に複数のアフィニティーチップに導入することが可能である。また、吸引と吐出後に試料溶液を攪拌させることで、原理的に吸引吐出量の数十倍の試料溶液からの精製が可能となる。またアフィニティーチップの作成が容易である点、吸着、洗浄、溶出の全工程の自動化が可能で

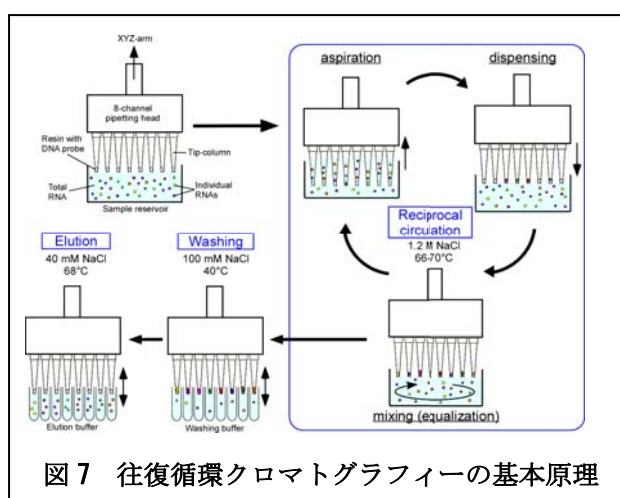


図 7 往復循環クロマトグラフィーの基本原則

ある点が特に優れている。さらに、マルチピペッターの本数を増やすことで同時に精製する検体数の拡張が容易である点が挙げられる。すでに我々は、8 検体用自動分注機をベースとし、DNA 固相化樹脂を詰めたチップカラムを用いることにより、全自動 RNA 精製装置の試作機を完成させている (図 8)。モデルを立て往復循環クロマトグラフィーの理論式を構築したところ、カラムに固相化したリガンド(DNA や抗体)と標的分子 (RNA やタンパク質) とのアフィニティー(平衡定数)により、

最終的な収率と十分な精製に必要な往復循環の回数を見積もることが可能である。一度に吸引する量やチップカラムの本数などを変化させた場合でも必要回数などの算出が容易である。このモデルの妥当性はすでに実験的に確かめられている。本試作機を用い、パン

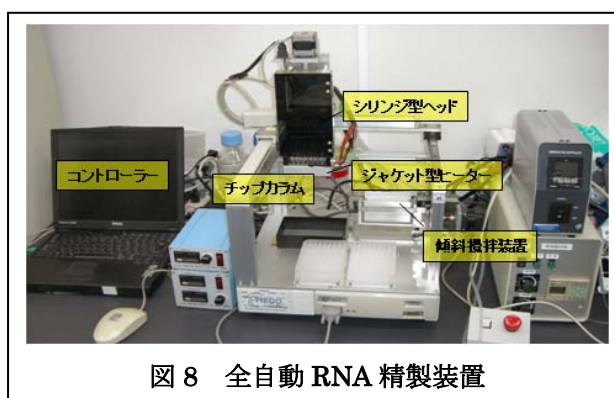


図 8 全自動 RNA 精製装置

酵母およびマウス肝臓由来の ncRNA を全自動で単離することに成功している (図 9)。また、miRNA や mRNA の単離にも成功しており、本装置を用いることで細胞内に存在するほぼすべての RNA が精製可能であると考えて

いる。さらに、抗体を固定化したアフィニティーチップを用いることにより、全自動で多種類のタンパク質を同時に免疫沈降することにも成功しており、今後は本装置を用いて、様々な ncRNA を単離解析することで、RNA の機能発現に必要な修飾構造を決定したり、機能未知な RNA 結合タンパク質を精製し、結合している RNA を同定していきたいと考えている。

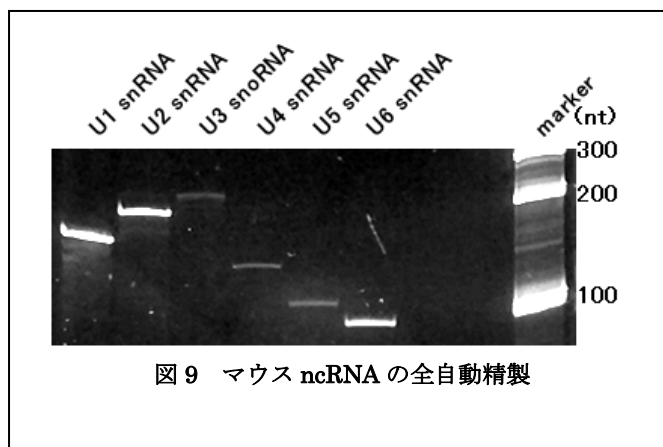


図 9 マウス ncRNA の全自動精製

現在、民間企業と世界初の全自動多検体アフィニティークロマト装置の商品化に向けた共同開発を行っている。

## 2. 2. 1. 6. ヒト脳トランスクリプトームにおけるイノシン修飾の網羅的同定

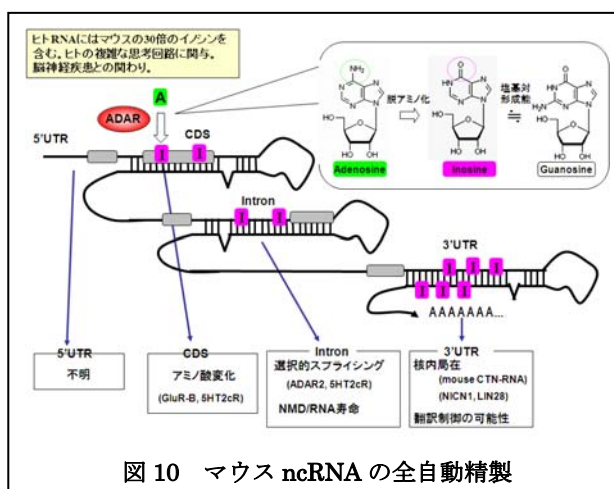
A-to-I RNA エディティングとは、二本鎖構造を形成した RNA 中のアデノシン(A)が脱アミノ化され、イノシン(I)へと修飾される機構である (図 10)。イノシンは化学構造上グアノシン(G)と類似しており、シトシン(C)との塩基対形成が可能である。そのため、イノシン化修飾により遺伝暗号や二次構造の変化が引き起こされる。この機構は二本鎖 RNA 特異的アデノシンデアミナーゼ

(ADAR=adenosine deaminase acting on RNA)

により行われ、哺乳動物には ADAR1、ADAR2、

ADAR3 の 3 種類の酵素が存在する。ADAR1 ノックアウトマウスは胚性致死であり、発生段階で肝臓、心臓、椎骨などで異常なアポトーシスが観測される。特に肝臓では血球分化の際にインターフェロン応答が誘導され、アポトーシスが引き起こされることが知られている。この表現系から ADAR1 によるイノシン化修飾が発生や分化に関与していると考えられている。ヒト RNA にはマウス RNA の 30 倍ものイノシンが含まれていることが知られており、ヒトの複雑な思考や神経回路の形成に関わっていることが指摘されている。遺伝性対側性色素異常症 (DSH、Dyschromatosis Symmetrica Hereditaria) の患者において ADAR1 のヘテロナンセンス及びミスセンス変異が発見されており、皮膚の色素斑/白斑、脳の石灰化、筋緊張異常や精神遅滞・機能低下等の症状が報告されている。グルタミン酸受容体のサブユニットである GluR-B の mRNA は ADAR2 の基質であり、エキソン 11 のグルタミンをコードする CAG コドン中の A がほぼ 100%の効率でイノシンに修飾され、アルギニンへとアミノ酸が変化する。このアミノ酸配列の変化がグルタミン酸受容体のカルシウム透過性をコントロールしているため、ADAR2 のノックアウトマウスではてんかん症状が起こり早期に死に至る。また、筋萎縮性側索硬化症 (ALS) 患者の運動神経細胞では、この部位のエディティング効率が顕著に低下していることが発見されている。

イノシン化部位は逆転写反応で cDNA に変換すると G として読まれるため、見掛け上、A から G へ配列が編集される。イノシン化部位を同定するために、従来法では、同一組織または細胞由来のゲノムと RNA をそれぞれ PCR と逆転写 PCR で増幅し、配列を比較することでイノシン化部位を特定している。しかし、この手法では偽遺伝子に由来するシグナルや、シーケンスエラーやノイズなどと識別するのが困難であるなどの問題がある。我々はイノシン塩基特異的な化学修飾と逆転写反応を組み合わせた生化学的なイノシン化部位特定法である ICE (inosine chemical erasing) 法を開発した。化学修飾剤としてアクリロニトリルを用い、イノシン塩基の 1 位を特異的にシアノエチル化することで、逆転写による cDNA の伸長をイノシン化部位の手前で止めることができる。したがって化学修飾の処理と未処

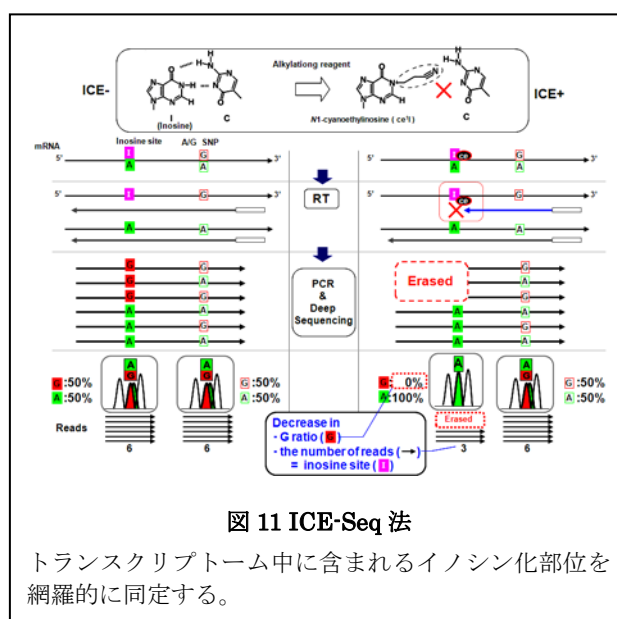


理で逆転写 PCR を行い cDNA を比較することで、イノシンに由来する G のシグナルを特異的に消失することが可能になる。この手法により、イノシン修飾を、アレル間の SNP やシーケンスエラーやノイズと明確に区別することができるようになった。また、ゲノムワイドなイノシン修飾の解析作業と同定部位の登録作業の効率化を図るため、各配列データのアライメント、イノシン化部位の正誤判定、イノシン化率の測定、データベースへの登録・閲覧を半自動的に行えるソフトウェア (ICE-CAFÉ) を開発した。

まずは、公開されている約500万のESTデータベースとヒトゲノム配列の比較から絞り込まれたA/G置換部位をA-to-Iエディティング候補部位とし、ICE法を用いたゲノムワイドな解析を行った。最終的に、約4,000領域の解析を実施し、約20,000箇所のイノシン化部位を特定した。解析した領域にはイスラエルのCompugenのグループが計算科学的な手法で予測したイノシン化部位が6,000箇所程度含まれていたが、このうちイノシンであることが確認できた箇所は3,800箇所程度 (57%) に留まった。残りの4割強の箇所については、EST上で見られるSNPやシーケンスの間違い、あるいは我々が解析した個体では観測されないイノシン化部位であると考えられる。一方で全体の80%に相当する16,000箇所に関しては完全に新規部位であり、我々のICE法による同定手法の優位性が示されたことになる。同定された部位の多くはmRNAの長鎖3' UTR内に存在するAlu反復配列上に存在した。ESTデータベースや各mRNAについての報告から、これらmRNAにはバリエーションが存在し、イノシン化部位領域を全く含まない短鎖3' UTRを持つものとイノシン化部位領域を含む長鎖3' UTRをもつものが共存することが考えられる。

さらに網羅的にイノシン化部位を特定する手法として、ICE法と次世代シーケンサー (Genome analyzer, Solexa) を組み合わせた手法 (ICE-Seq) を考案した (図11)。

解析対象のRNAをイノシン特異的なシアノエチル化の処理 (ICE-) と未処理 (ICE+あるいはICE++) で調製し、cDNAを大量シーケンスを行い、全遺伝子に貼り付けた後で、ICE-の条件でA/G置換部位を検出する。イノシン化部位はシアノエチル化後 (ICE+あるいはICE++) でイノシンに由来するGのリードが特異的に減少するため、情報処理を行うことでイノシン化部位の検出が可能である。ヒト成人脳由来のポリA+RNAをシアノエチル化の処理と未処理で調製し、mRNA-Seqのプロトコルで約300塩基対のcDNAを合成した。Genome analyzerを用い、75塩基長のペア-エンドで最終的に27レ



グソフトウェアにBWA5.1を使用し、UCSC geneをリファレンス配列として各リードを貼り付けたところ、全体の約6割のリードを貼り付けることができた。結果として、ICE-が173億塩基、ICE+が156億塩基、ICE++（強条件）が171億塩基分のデータが得られた。遺伝子に貼り付けたリードの平均重複度が20X以上のものを解析対象とした。Gのリードの特異的な減少を指標にイノシン化部位を絞り込んだところ、現時点で約18,000箇所のイノシン化候補部位を絞り込むことに成功した。実際この中に、我々が先に特定したイノシン化部位が2,300箇所含まれていた。また、コーディング配列内には、約80か所のイノシン化部位が見出されたが、実際この中に、既報の19箇所が含まれていることから、ICE-Seqの同定精度の高さが窺える。最終的には、約30,000箇所の新規イノシン化部位を報告できる見込みである。

## 2. 2. 1. 7. まとめ

### (1) 実施計画書記載の達成目標に対する到達度

RNA-MSの開発により、機能性RNAを高感度(サブフェムトモルオーダー)で定量的かつ直接測定する手法を開発する、という本プロジェクトの最終目標を十分に達成した。また、RNA-MSを実用化するために、試料の前処理システムや情報処理技術の開発にも成功し、国際特許の出願や著作物を登録しており、知財戦略の面でも優位な状況にある。

### (2) 実用化・事業化の見通しについて

RNA-MSは機能性RNAの基礎研究のみならず、核酸医薬の品質管理や薬物動態にも十分に応用可能な技術として応用可能な手法である。RNA-MSデコンボリューション（東大著作物登録）、CIDチェッカー（国際出願）、RNAマスフィンガープリント法（国内出願）、往復循環クロマトグラフィー（国際出願）など主要な関連技術を確保しており、国内の企業を中心に技術移転を進めていく予定である。特にMSメーカー、製薬企業、臨床試験受託機関との連携を予定している。

## 2. 2. 2 機能性RNAの高感度検出システムの開発

集中研②、分室5（ノバスジーン/オリンパス）、分室6（DNAチップ研究所）

共同実施先：産総研②、東京大学（3；陶山）、北陸先端科学技術大学院大学（藤本）  
武庫川女子大（村田）

### 2. 2. 2. 1. 光反応を利用した Photo-DEAN (DNA-Encoding-based ANalysis) 法および 1分子蛍光分析法による機能性RNA高感度検出システムの開発

集中研②、分室5（ノバスジーン/オリンパス）、

共同実施先：産総研（2）、東京大学（3）、北海道大学、  
北陸先端科学技術大学院大学

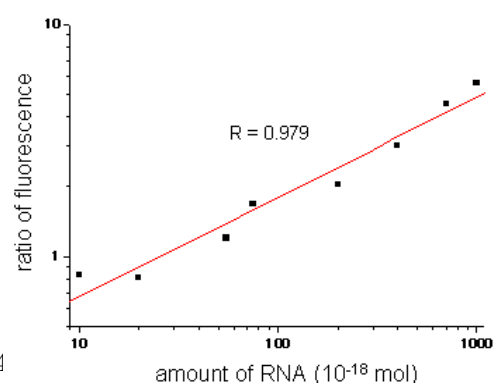
機能性RNAにはmicroRNAに代表される短鎖RNAやHIF1 $\alpha$ のNatural antisense RNA (NAT)、またmRNA型ncRNAなど様々な種類が存在している。よく研究されているmicroRNAの解析には、アレイや定量的PCRによる解析方法が開発されているが、これらの方法はmicroRNAに特化したものであり、microRNAとその他のRNAの定量解析する場合、各々異なった手法を用いる必要がある。そこで、本プロジェクトの開始以前から取り組んでいたcDNA高感度定量法 (DEAN法) によるmRNA定量検出方法を応用することにより、その鎖長やpolyA鎖の有無に関わらず同じ操作で定量開発する方法の開発を目指した。DEAN法による遺伝子発現解析はpolyA RNAをcDNAに変換したのち、標的cDNA配列をハイブリダイゼーションおよびライゲーションにより特性のよくわかった既知配列に置換し、増幅・標識を行い、発現解析を行うものである（特許取得済み）。そのためDEAN法ではPolyAを持たないRNAの解析は難しかった。そこで、藤本教授（北陸先端科学技術大学院大学）らにより開発された光ライゲーション技術を応用し、機能性RNAをダイレクトに既知配列へ置換することにより、機能性RNAの発現解析技術を可能とするPhoto-DEAN法の開発を行った。また、microRNAは血清中に intact な状態で存在することが知られる一方で、がんと正常細胞では発現レベルが異なっていることなどが近年の研究成果で判明しており、新たなバイオマーカー候補としての期待が高まっている。そこで、将来の臨床応用を視野にいたし、「非増幅でmicroRNAを定量検出するアッセイ系」の開発を行うことにより、非増幅atto molレベルの検出感度を達成することができた。

### 2. 2. 2. 1. 1. Photo-DEAN法による機能性RNA検出技術の検討

東大陶山研究室で開発されたDNAコンピューティングを活用したcDNA高感度定量法

(DEAN法)を、non-coding RNAに適用する研究を行った。DEAN法では、初期プロセスに標的cDNAを鋳型とする酵素ライゲーション反応が含まれるが、RNAを鋳型とする酵素ライゲーションは非常に効率が低く実用的ではない。そ

合成RNAサンプル量と蛍光強度比の相関



2.2.14

図1. Photo-DEAN法によるRNAの定量



ここで北陸先端大藤本研究室で開発された、RNA を鋳型とする DNA 連結が可能な光ライゲーション技術を応用した (photo-DEAN 法)。

Photo-DEAN 法により、8 種類の合成 RNA (30mer) の混合サンプルに対して個別に定量することに成功しており (図 1. 参照)、検出下限  $1\text{amol}$  ( $10^{-18}\text{mol}$ ) の感度を達成した。また、光ライゲーション法の導入により、ライゲーションプロセスは数時間からわずか 5 分に迅速化された。DEAN 法は、現在のラボスケールで、600 種類の RNA を 2 チューブ・1 チップで同時に処理できるシステムなので、RNA ごとに個別の検量線が必要な qRT-PCR 法と比べて、photo-DEAN 法では、2 桁高いハイスループット性が原理的に実現可能である。

## 2. 2. 2. 1. 2. Photo-DEAN 法によるマイクロ RNA の検出技術の検討

Photo-DEAN 法を用いた miRNA の測定系を開発するため、まず対象とする miRNA の選抜を行った。miRNA は 20~23 塩基と非常に短いため、測定可能な RNA は慎重に選抜する必要がある。Photo-DEAN 法に用いている光ライゲーション反応にはカルボキシビニルウリジンを用いており、連結相手はピリミジンである必要がある。そのため、測定する RNA 鎖にはプリンとアデニンが隣り合った配列、つまり、GA もしくは AA を内部に持っている必要が生じる。そこで、30 種類の miRNA を選抜してプローブの設計、合成を行った。本検討では、Photo-DEAN 法の miRNA への適合性を検証する目的で研究をおこなったため、カルボキシビニルウリジンのみを用いたが、カルボキシビニルシトシン、カルボキシビニルアデニンの合成にも成功しており、GG、AG、GU、AU を持つ miRNA に測定対象を広げる事も可能である。

選抜した 30 種類の miRNA を用いて Photo-DEAN 法を行った。以下にグラフを示すが、miRNA の種類に関わらず、RNA 量が  $100\text{amol}$  の時にシグナルが 1、 $1000\text{amol}$  の時にシグナルが 10 になる直線上に乗る状態が理想となるようにシステムを構築した。miRNA の違いによる上下のばらつきや傾きによって測定が上手く行っているかどうかを判断できる。MicroRNA の濃度を 20、100、 $500\text{amol}$  と量を変化させて測定した結果、投入量に比例してシグナルが得られるものと、シグナルが低く、存在量に合わないものにグループ化できることが分かった。これらの配列を詳細に解析したところ、量比を反映しなくなっているものは、内部に小さなヘアピン構造を作るものが多い事が分かった。T<sub>m</sub> を計算したところ、二重鎖の T<sub>m</sub> が十分に分子内ヘアピンの T<sub>m</sub> を上回っている場合はシグナルが得られるが、差が十分でない、もしくは、逆転しているような場合では量比を反映しなくなると考えられた。

これらの存在量と合わない miRNA の定量を改善するため、ハイブリダイゼーションの温度や反応液の塩濃度、およびデコイとなる核酸の影響について検討を行った。

反応温度を  $30^{\circ}\text{C}$  から、 $4^{\circ}\text{C}$  に変更する事で miR-96、miR-137、miR-141 において、存在量に応じたシグナルの変化が見られるようになり、定量性が改善したと考えられた (図 2)。

塩濃度は  $50\sim 150\text{mM}$  までは大きな変化は無かったが、 $200\text{mM}$  を越えると一気に定量性がなくなる事が分かった (図 3)。グラフから上下の差が大きくなり、傾きも緩くなっているのが分かる。

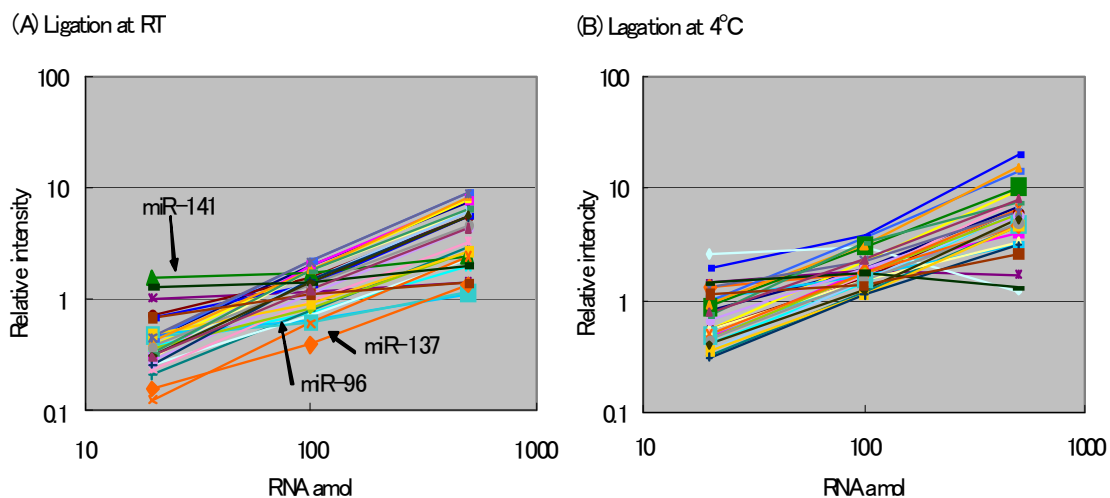


図2. Photo-Ligationの温度の影響。  
Ligation 温度が30°C (A) に比べて、4°Cで行うとmiR-96、137、141で定量性が向上している。

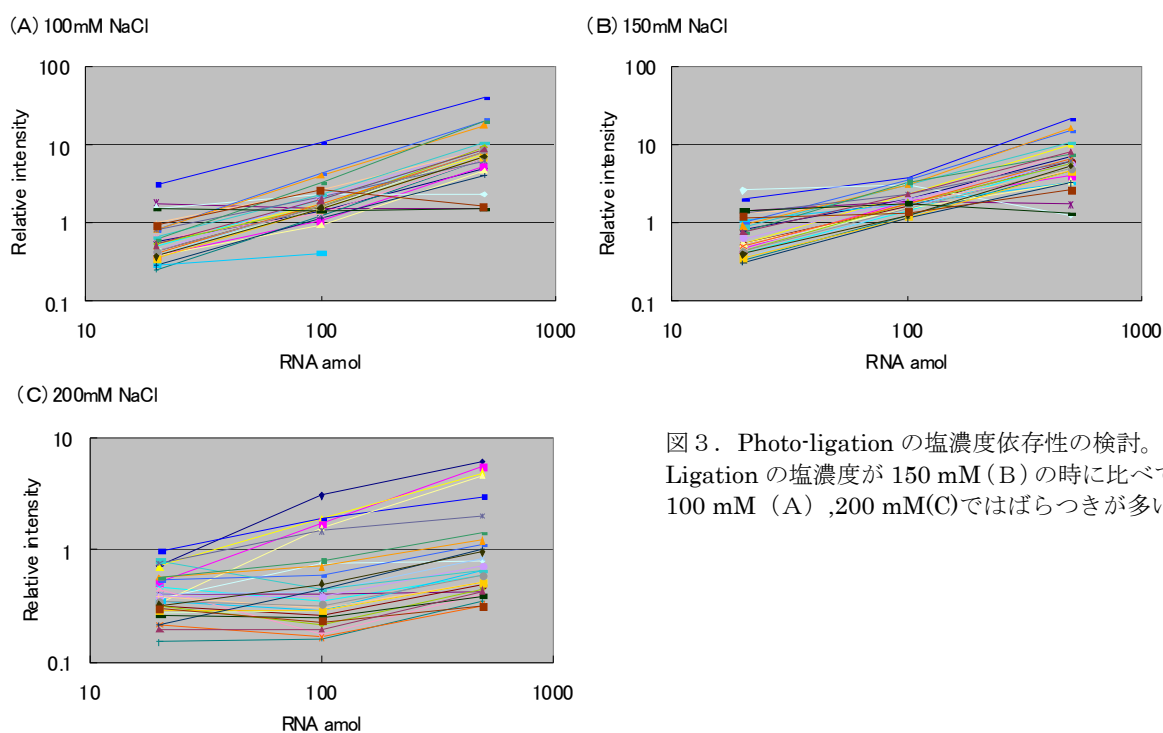


図3. Photo-ligationの塩濃度依存性の検討。  
Ligationの塩濃度が150mM (B)の時に比べて、100mM (A)、200mM (C)ではばらつきが多い

さらに、非特異的な吸着の影響を考慮し、デコイとなるRNAおよびDNAを混合して実験を行った。測定対象のmiRNAとの干渉を避けるため、デコイRNAには大腸菌のtotal RNAを用い、デコイDNAには規格化配列とは異なり干渉しないことが分かっている正規直行配列を元に合成したオリゴDNAを用いた。その結果、miRNA間のバラつきが多くのmiRNAで減少した。さらに、酵母由来の標準RNAによる検量線ではガタつきが減少し、測定精度が向上した。

次いで、実サンプルの検証として、HeLa 培養細胞から抽出したトータル RNA を用いて検出実験を行った (図 4)。

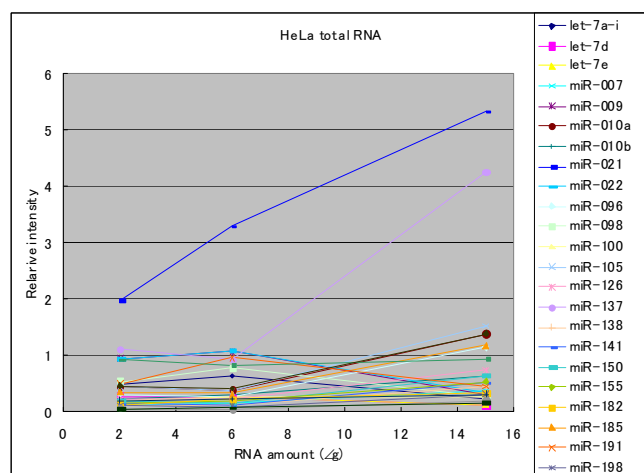


図 4. Hela total RNA を用いた Photo DEAN 法。miR-21 が投与した RNA に応じた変化を示した

その結果、高発現している報告の有る miR-21 について投入 RNA を変化させたときにそれに応じたシグナルが検出された。この事から、様々な RNA が存在する中で、実際に miRNA が検出可能で有る事が分かった。

また、選抜した miRNA の一部で FCS によるライゲーション効率の測定を行った (図 5)。

その結果、13 種類中、11 種類において十分な拡散時間の増加が見られ、約 20~60%の効率を示した。これは、Photo-DEAN 法による定量目的において十分な効率を示していると考えられる。一方、2 種類のものでは数%程度しか反応が進行していないものがあり、反応効率の向上が必要となるものが存在した。特に miR-16 については上記の DEAN 法測定においても結果が安定しない場合が多かったが、その原因としてライゲーション効率が低い事が考えられた。

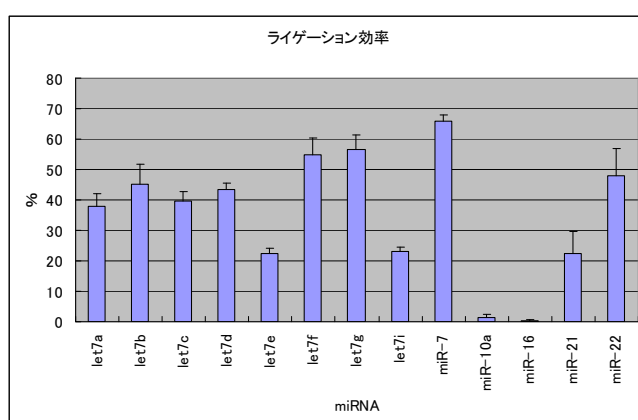


図 5. FCS による miRNA を鋳型とした光ライゲーションの効率

### 2. 2. 2. 1. 3. RNAをターゲットとした光ライゲーション反応の基礎的性質の検討（蛍光相関分光法を用いて）

上述のように、photo-DEAN 法を用いた RNA の発現解析技術では、miRNA のような短い RNA から polyA タイプの機能性 RNA や mRNA ような比較的長い RNA を同一のプラットフォームで定量解析することが可能となる。そのため、解析のハイスループット化や、同一プラットフォームによるデータの整合性等、有益な情報が得られることが期待できる。PolyA RNA のような長鎖 RNA では、特異的なプローブ設計が可能であり photo-DEAN 法により attomol レベルの検出感度が得られている。しかし、miRNA のように 22mer 程度の短い RNA では、各プローブの鎖長は 10mer 程度であり、その効率と特異性を得るための条件検討が重要である。そこで、特に特異性を得ることが難しいと想定される let7 ファミリーについて蛍光相関分光法を用いて、そのライゲーション効率と特異性について検証した。検証用のターゲットは合成 RNA とし、let7a をターゲットとして、ライゲーション効率を検証した。アンカープローブ（蛍光プローブ）は let7a 配列の 3' 側 10mer の相補的 DNA 配列に Tamra 標識したものをを用い、アダプタープローブとして let7a の 5' 側 12mer に 69mer のコード配列を付加したものをを用いた。この両プローブが連結した場合、蛍光プローブのみの 10mer が、91mer となり拡散時間が増大することで連結したプローブの割合が求められる。ライゲーション温度を 10°C、5 分間光照射を行った結果、約 70% の蛍光プローブがライゲーションしていた。さらに、特異性を検証するために、let7b の合成 RNA をターゲットとし、let7a と同一条件で光照射を行った場合、let7a をターゲットとした場合と同程度のライゲーション効率であり特異性は得られなかった。そこで、ライゲーション温度を 10°C から 30°C まで変化させたところ、30°C で ligation した場合、特異的な let7a に対するライゲーション効率はほとんど変化せず、let7b をターゲットにした非特異的ライゲーションのみ減少した（図 6.A）。さらに、ホルムアミドを 2.5% 添加することにより、非特異反応の割合は 30% 程度となり特異性の向上がみられた（図 6.B）。

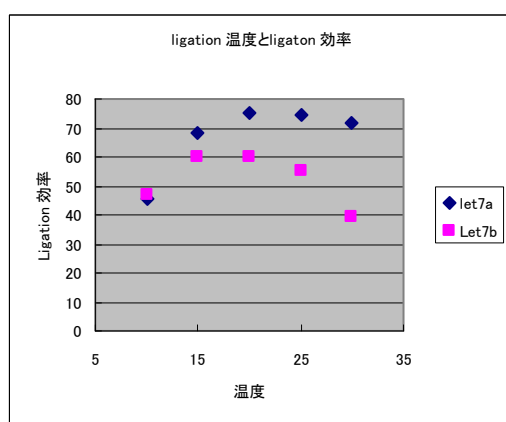


図 6.A ライゲーション温度と特異性

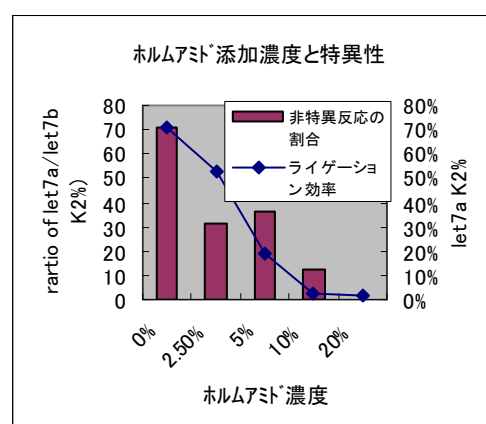


図 6.Bホルムアミド濃度と特異性

## 2. 2. 2. 1. 4. カルバゾール類縁体を用いた光可逆的な RNA クロスリンク反応開発

上述のように H19 年度までに、我々は光ライゲーショと蛍光相関分光法あるいは DEAN 法を組み合わせ miRNA の高感度定量検出方法を検討してきた。光ライゲーショ反応は、ピリミジン塩基が隣あっている必要があり、成熟 miRNA のような短鎖 RNA ではライゲーショポイントが限定され、特異的な検出の難しい場合があった。そこで、H20 年度からは、ターゲットの miRNA と検出プローブを結合させる新規な光クロスリンク反応を開発し、非増幅な miRNA 検出方法の開発に着手した。

光クロスリンク反応とは、光応答性塩基を含む ODN (oligodeoxynucleotides) と一本鎖 RNA をハイブリダイゼーションした後に、光照射することで光応答性塩基と RNA に含まれる塩基との間で光環化反応が進行して、塩基間を共有結合でつなぐ反応である。これまでに、DNA との光クロスリンク反応は数多く報告されているが、RNA との光クロスリンク反応はほとんど報告例がない。そこで、新しい光応答性塩基としてシアノビニルカルバゾールヌクレオシド (<sup>CNV</sup>K) を合成して、その光クロスリンク反応を検討した (図 7)。<sup>CNV</sup>K を含む ODN と RNA をハイブリダイゼーションした後に、366nm 光を 1 秒間照射したところ、<sup>CNV</sup>K を含む ODN と RNA のピークが消失して、新しいピークとして光クロスリンク体が 97% の収率で得られた (図 8 A)。一方、その光クロスリンク体に 312 nm 光を 60 秒間照射したところ、それらのピークは消失して、元となる <sup>CNV</sup>K を含む ODN と RNA のピークが 97% の収率で生成した (図 8 B)。

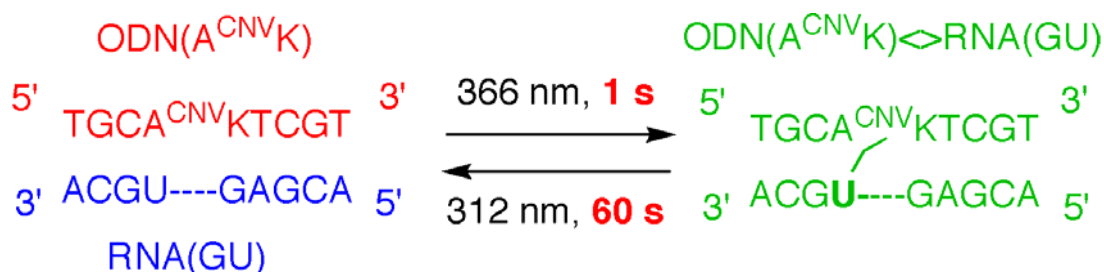


図 7. ODN(赤)中の CNVK は対合する塩基の 1 塩基 3'側のピリミジン塩基と光結合する (緑)。

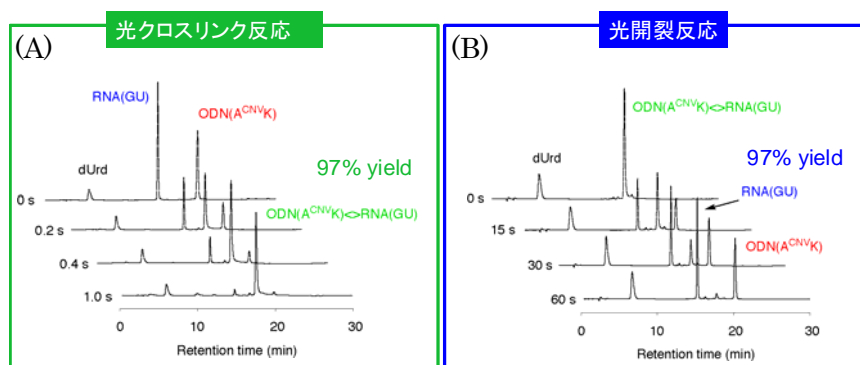


図 8. <sup>CNV</sup>K を含む ODN と RNA をハイブリダイゼーションし、366 nm の光を 1 秒間照射した。照射時間毎のサンプルを HPLC により解析したところ、1 秒で 97% の ODN と RNA のクロスリンクを観察した (A)。さらに 312 nm の光を 60 秒間照射すると 97% の効率でクロスリンクの解裂が観察された (B)。

## 2. 2. 2. 1. 5. 可逆的 RNA 光クロスリンク反応を用いた配列選択的 RNA 選別法の開発

図9は光クロスリンク反応を用いた配列選択的 RNA セレクションの模式図である。光応答性塩基である <sup>CNV</sup>K を持つ光応答性プローブを標的となる RNA が含まれる試料液に入れて、366nm の光を照射すると標的 RNA に対してのみ光クロスリンク反応が進行する。この段階で RNA のセレクションが行われる。次に、固相担体からなる光クロスリンク体を試料液から取り出すことで RNA が精製される。取り出した光クロスリンク体に 312 nm 光を照射することで、標的 RNA のみが回収されることになる。

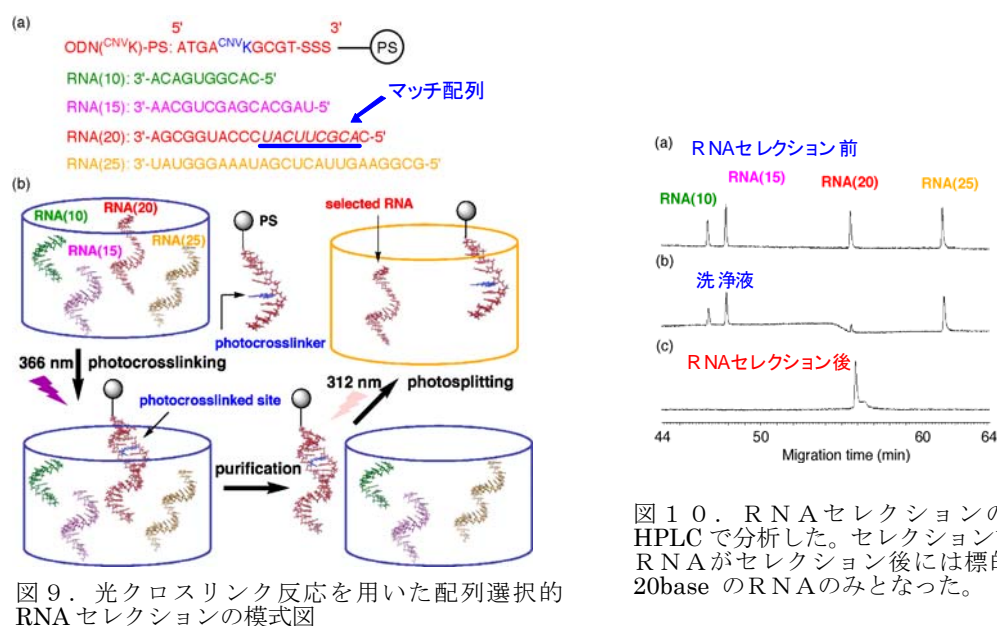


図10. RNAセレクションの過程をHPLCで分析した。セレクション前4本のRNAがセレクション後には標的とする20baseのRNAのみとなった。

実際に、RNAセレクションの過程をキャピラリーゲル電気泳動で分析したところ、セレクション前は4本のRNAが存在するのに対して、洗浄することで3本のRNAに減少して、有効であることが示された。

さらに、我々は本方法により、複数のmiRNAの中から目的のmiRNAを配列選択的に選別することに成功した。図10は光クロスリンク反応を用いた配列選択的 RNA セレクションの模式図である。光応答性塩基である <sup>CNV</sup>K を持つ光応答性プローブを標的となる miRNA が含まれる試料液に入れて、366 nm の光を照射すると標的となる miRNA (let 7a) に対してのみ光クロスリンク反応が進行した。この段階で RNA のセレクションが行われる。次に、固相担体からなる光クロスリンク体を試料液から取り出すことで RNA が精製され、取り出した光クロスリンク体に 312 nm 光を照射することで、標的 RNA (let 7a) のみが回収されることになる。実際に、RNA セレクションの過程をキャピラリーゲル電気泳動で分析したところセレクション前は3本の RNA が存在するのに対して、洗浄することで2本の RNA に減少して、

本手法が選択的な microRNA 選別法に有効であることを見出した (図 1 1)。これらの結果について ChemBioChem, 2009, 10, 1473 に論文としてまとめた。

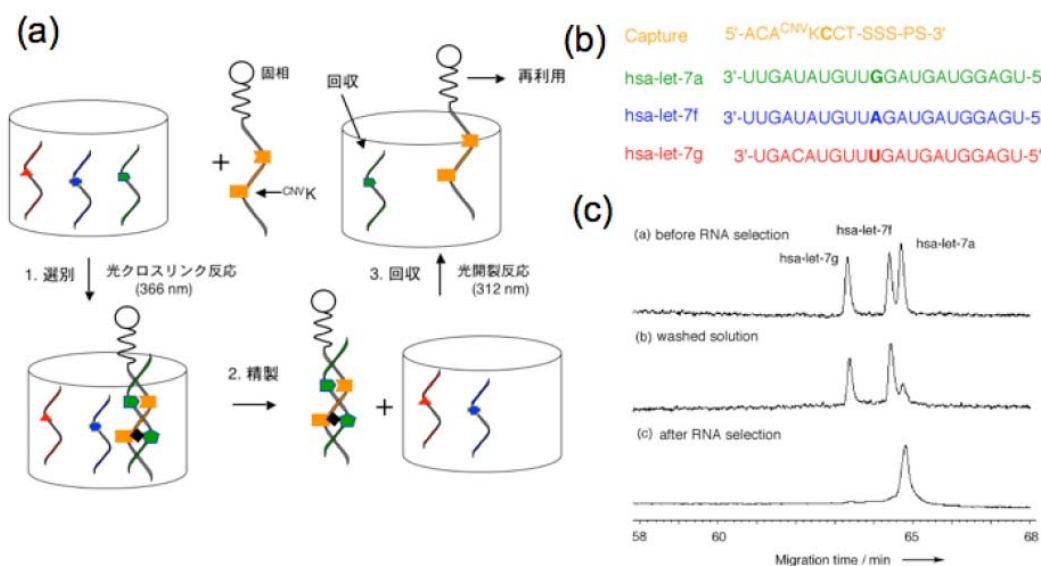


図 1 1 . . (a) Strategy for selection of a target RNA sequence. (b) Capture and miRNA sequences used in this study. Italic characters indicate a matched sequence with a capture strand. Bold characters indicate a single mismatch. (c) CGE analysis for each operation: (a) before RNA selection; (b) the washed solution after irradiation at 366 nm; (c) after RNA selection.

本方法を利用することでマイクロ RNA のようなノンコーディング RNA といった翻訳を受けない RNA を網羅的にスクリーニングする方法として期待できる。さらに遺伝子クローニングでは、特定の RNA を細胞内から精製した後に RNA を転写する方法で行われている。本方法を応用することで、高い回収率で RNA を精製できるため、標的の遺伝子をクローニングする際に有用と思われる。

## 2. 2. 2. 1. 6. 1 分子蛍光法による非増幅高感度 miRNA 定量方法の開発

### (1) 蛍光共鳴エネルギー移動 (FRET) を利用した FRET-FIDA 法の検討

H 1 9 年度まで我々は、光ライゲーション反応と蛍光相関分光法を組み合わせ、高感度な miRNA を定量する技術を開発してきた。光ライゲーション反応はピリミジンが隣あった配列であることが必要であり、miRNA のような短い RNA をターゲットにした場合、1 部の miRNA では特異性をもったプローブを設計できないという問題があった。また、蛍光相関分光法では実際の測定時間 (30 秒程度) では femto M オーダー (sub nM) の検出能であり、より高感度な測定が望まれた。そこで、より高感度に測定できる 1 分子蛍光法として、蛍光

強度分布解析法 (FIDA 法 Fluorescence Intensity Distribution Analysis:FIDA) と蛍光共鳴エネルギー移動 (Fluorescence Resonance Energy Transfer, FRET) を組み合わせることで高感度な測定系の開発を行った。蛍光相関分光法は、共焦点領域内に入出入りする蛍光の時間による揺らぎをもとに解析する。一方 FIDA 法は測定溶液中に 2 種類以上の異なる蛍光強度をもった分子が存在する場合に B/F 分離をせず、蛍光分子の蛍光強度と分子数を計測する方法である。本方法を用いると、レーザー照射領域を回転することにより、FCS より希薄な溶液であっても、短時間で測定できる。そこで、FIDA 法を miRNA の測定に応用するために FRET と組み合わせ、非増幅で miRNA を定量できる新規な方法の開発 (FRET-FIDA 法) を行った。

図 1 2 は miR-21 と相同な配列をもつ合成 DNA を本方法により検出した例である。マイクロ RNA、miR-21 と相同な配列をもつ 1 本鎖 DNA を 0 attomol ~ 5 fmol/ $\mu$ L の範囲で添加し、検出限界と定量性を検証した。その結果少なくとも 250 amol/50  $\mu$ L の検出感度があることがわかった。また 1 fmol/ $\mu$ L でも飽和していないことから 1nM のプローブ濃度で、atto mol から femto mol まで少なくとも  $10^3$  のダイナミックレンジで測定可能である。今回の検討例はプローブ濃度を 1nM 程度としたが、プローブ濃度をさらに低濃度とすることで、検出感度をより向上させることも可能であると考えられる。本アッセイ法に関しては特許出願を行った。

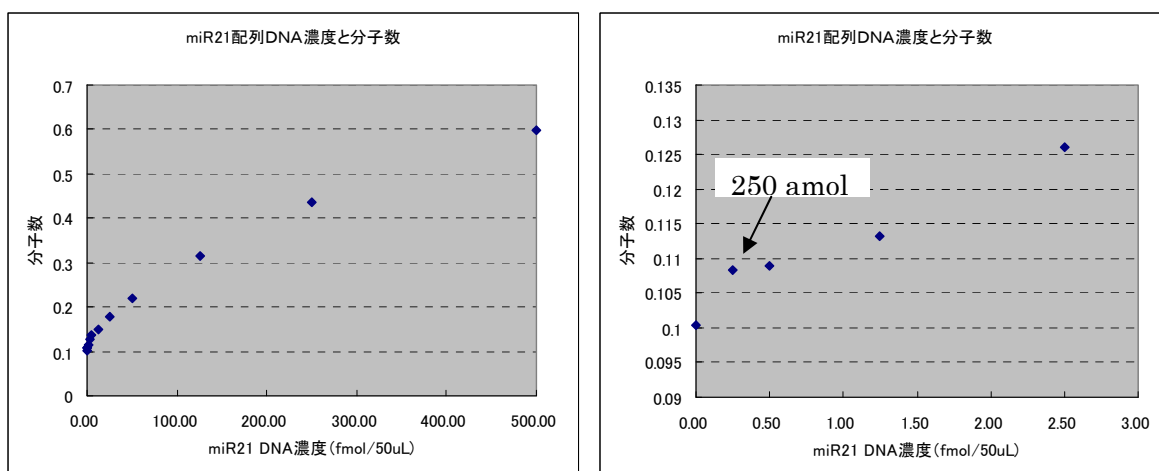


図 1 2. FRET と FIDA を組み合わせた高感度 miRNA 検出方法における測定例。miR21 と相同な配列をもつ合成 1 本鎖 DNA を本法で検出した。MF20 による測定では、1 ウェルあたり 50  $\mu$ L 程度の溶液が必要であり、x 軸は 50  $\mu$ L に含まれる miR21 配列 DNA 量を表している (250 attomol のとき濃度は 5 pM となる)。測定系を工夫することで 10  $\mu$ L 以下のサンプル量にすることも可能であるので、50 attomol 程度の検出も可能である。

## (2) FRET-FIDA 法にクロスリンクを組み合わせた高感度検出方法の検討

FRET-FIDA 法は 250 attomol/50  $\mu$ L の miRNA を非増幅で検出できるシステムであるが、1 分子蛍光法は室温測定であるため、室温中での蛍光標識 probe とターゲット RNA の平衡反応となるため、サンプル中に存在するターゲット RNA の類似配列 RNA との非特異的



なハイブリダイゼーションを防ぐことが難しい。そこで、本 FRET-FIDA 法に北陸先端大で開発された光クロスリンクを組み合わせたことによる特異性の向上について検討を行った。光クロスリンク反応を用いることにより、特異性の高い温度あるいは溶液条件でプローブとターゲット miRNA のハイブリダイゼーションとクロスリンクが行われることで非平衡な状態となり、室温測定でもストリンジェンシーの高い環境が保持され、特異性高く測定できる。

Tamra 標識プローブに光クロスリンク剤である <sup>cnv</sup>K を 1～2 箇所を導入したプローブを用い、光クロスリンクによるクロスリンクを行ったのち、終濃度 50% となるようにホルムアミドを添加し、クロスリンクしていないプローブを解離させることでそのクロスリンクの効率を検討した。

図 1 3 は検討結果である。X 軸は光照射の有無 (+/-) とホルムアミドの添加の有無 (+/-) である。本検討では、ターゲット miRNA とハイブリダイゼーションをした時にプローブの蛍光が消光するようにプローブを設定している。そのため Y 軸は、ターゲット miRNA とハイブリダイゼーションしていないプローブ数をあらわしている。X 軸の右端の光照射無し、ホルムアミド添加では、全てのプローブで 30 以上の蛍光プローブがハイブリダイゼーションしていないのに比べて、光照射後では、<sup>cnv</sup>K を含んでいないプローブのみが、

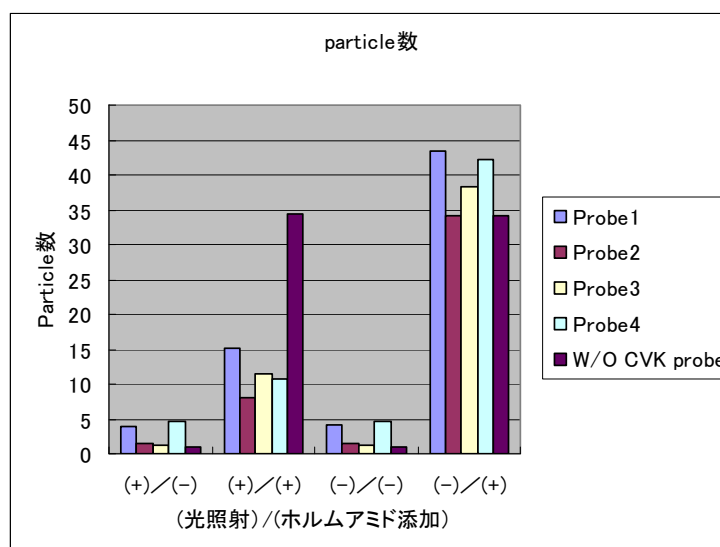


図 1 3. <sup>cnv</sup>K の位置と光クロスリンクの効率

突出してハイブリダイゼーションをしていないフリーな状態で計測され、クロスリンクは有効に作用していることがわかった。Probe-1 は <sup>cnv</sup>K の導入位置が、中央であるため、ハイブリダイゼーション効率が低く、クロスリンクしているプローブ数も少ないものと考えられた。そこで最もクロスリンク効率の高い Probe-2 を用いて miR-21 配列と相同な合成 RNA を用いて感度を検討した。プローブ濃度は 1nM とし、添加 miR-21RNA 濃度を 500fmol/50  $\mu$ L までとした。その結果、5 fmol/50  $\mu$ L の検出感度であった。

FRET-FIDA 法に光クロスリンクを組み合わせることで特異性が高くなる可能性が見えた反面、感度では1桁程度悪くなっていた。この原因を詳細に分析した結果、蛍光 probe 内での鎖内2次構造が影響している可能性が考えられた。そこでさらに適合性について検証を行い、pM から nM の範囲で線形近似で決定係数 ( $R^2$ ) が 0.9985 となる直線性を示し、またさらにその検出限界が、15 pM から 30 pM であるようなアッセイ系を開発することができた。これは、1  $\mu$ l 換算で 30 attomol の測定も可能であると考えられた。さらに特異性を伴って検証するために、miR-21 配列を含み全 8 種類の合成 RNA をターゲットとし、測定したところ、miR-21 のみに特異的に測定することができた。

### 2. 2. 2. 1. 7. まとめ

#### (1) 実施計画書記載の達成目標に対する到達度

1 分子蛍光法と北陸先端大 藤本研により開発された応答性ヌクレオシドとして 3-シアノビニルカルbazool 分子 ( $^{CNVK}$ ) を組み合わせることで、15 pM ~ 30 pM の miRNA 検出感度を達成した。本濃度は 15 attomol/ $\mu$ l に相当する。現状の測定装置では 384 well plate で測定するため、30  $\mu$ l の測定容量が必要であるため、450 attomol レベルであるが、1 分子蛍光法は原理的に 1  $\mu$ l 以下のサンプル容量で測定可能であるため、装置改良により 10 attomol 以下の miRNA の測定も可能であり、達成目標に到達している。Micro RNA は血清中で pM の程度の濃度で存在する可能性が指摘されており、本測定系は血清中の miRNA 測定が可能なレベルに到達したと考えられる。北陸先端科学技術大学院大学では昨年開発した光クロスリンク反応を用いた miRNA の選択的回収法の開発に成功した。

また、Photo DEAN 法による機能性 RNA 検出も attomol レベルの検出感度に到達しており、マルチプレックスに血清中の miRNA を測定することができるレベルに到達している。

#### (2) 実用化・事業化の見通しについて

マイクロ RNA はがんの制御に関わっていることが知られており、がんのバイオマーカーとして注目されている。また、血清 1  $\mu$ l あたり 10,000~100,000 個程度あるといわれており、これは、pM レベルの濃度で血清中に存在していることになる。これらのマイクロ RNA の高感度解析は、RNA を DNA に変換した後、定量的 PCR により定量している。本研究成果では、これらのマイクロ RNA を日本発の技術である光クロスリンク反応とオリンパスの保持技術である 1 分子蛍光法により、非増幅で検出できることを示したものである。

今後がんのバイオマーカーとしてマイクロ RNA が確立されることにより、本技術を応用した、がんの転移診断、予後診断、あるいは早期発見に応用可能と考えられる。

## 2. 2. 2. 2. MPEX法を用いたRNAの微量検出法の構築 DNAチップ研、武庫川女子大

RNAの検出技術に関しては、マイクロアレイに代表される集積型の発現量解析技術と、PCR増幅反応を応用したリアルタイム検出技術が確立されている。しかし近年話題となっている非常に短い機能性RNA分子を検出するためには、様々な技術改良が必要であった。本研究開発の開始当時におけるマイクロアレイ技術では、検出感度が通常のmRNAの検出感度と比較して数百～千倍程度も悪く、本研究開発中に発表されたアジレント社のアレイでは感度の問題は解消されたものの、精製したマイクロRNA(miRNA)を解析に使用できないなど説明不可能な現象も生じている(本論参照)。PCR技術ではABI社のTaqMan miRNA検出キットが毎年のように改良を重ね、初期のころの多種類同時検出の不便性を解消してきた。しかし同時に数百種類のmiRNAを検出するためには、数百万円にもなる初期投資が必要であり、これはアレイの実験でも同様である。

DNAチップ研究所では、住友ベークライトと共同開発により、アレイ上で通常の酵素反応を行える表面コーティングの開発に成功していた。このアレイコーティングを使用することにより、アレイ表面に固定したオリゴDNAをプライマーとしてDNAの伸長反応を行うことが可能であり、MPEX(Multiple Primer Extension)反応として特許を取得した。この時点では、RNAを使用する際には一度チューブ内で逆転写反応を行った後に、DNAをアレイ上で増幅・標識することを想定しており、RNAウイルスであるノロウイルスの検出法などを確立した。本研究開発では、この技術をmiRNAに応用するあたり、miRNA分子の短さからくる逆転写反応のバイアスなどを除くために、アレイ上に直接miRNA溶液を乗せて逆転写反応を行う実験系が可能かどうかという検証から始めることとした。もしこの反応が可能であれば、これまでのマイクロアレイの常識を覆し、通常チューブ内の酵素反応と同様に実験による個人差がほとんど無く、アレイ本体および検出試薬の値段も安く、一般の研究者に受け入れられ得るツールを開発できる可能性があると考え、本研究開発を開始した。

### 2. 2. 2. 2. 1. 目標

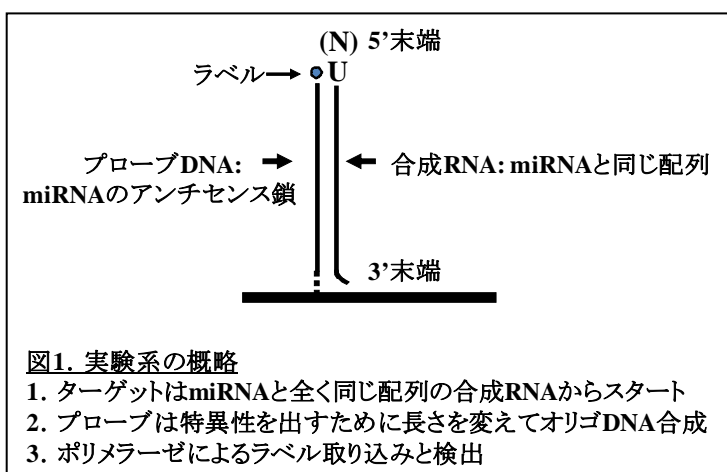
本研究開発では機能性RNAの発現解析を簡便にできるツールを開発した。機能性RNAの中でも50塩基以上のRNA分子に関しては、既存のマイクロアレイ技術や定量的PCR実験等の手法で十分解析可能である。そのため我々は解析が非常に難しい低分子機能性RNA、特にmiRNAをターゲットとして、マイクロアレイ上で多くの分子を同時に解析できる実験系の開発を行った。低分子RNAでは、①標識効率を上げること、②検出用プローブの自由度を上げて配列特異的認識を行うこと、③ハイブリダイゼーション(アニール)の効率を上げること、の大きく3点が問題になると考えられるが、本実験法の原理では全てを簡単な実験法で解決することができる。上記3点をクリアする際の目標として、

1. 検出限界の数値目標：1～数attoモル(=10<sup>6</sup>分子)レベル
2. 網羅性と配列特異性の両立(・バランス)
3. 実験手法の簡便性(とそれに伴うデータの安定性)

の3項目を挙げた。これらの項目を全て達成することができれば、先行している外国メーカーのツールに対抗することが可能になるだけでなく、マイクロアレイ実験に精通していない研究者に於いても、通常の酵素反応等と同様に一般的な実験法として普及する可能性が大きい。様々な高次生命現象過程を研究している研究者が、手元にあるサンプル群を使用して各々の分子マーカーを探索する国産ツールとして、非常に大きな役割を担えるものと考えている。

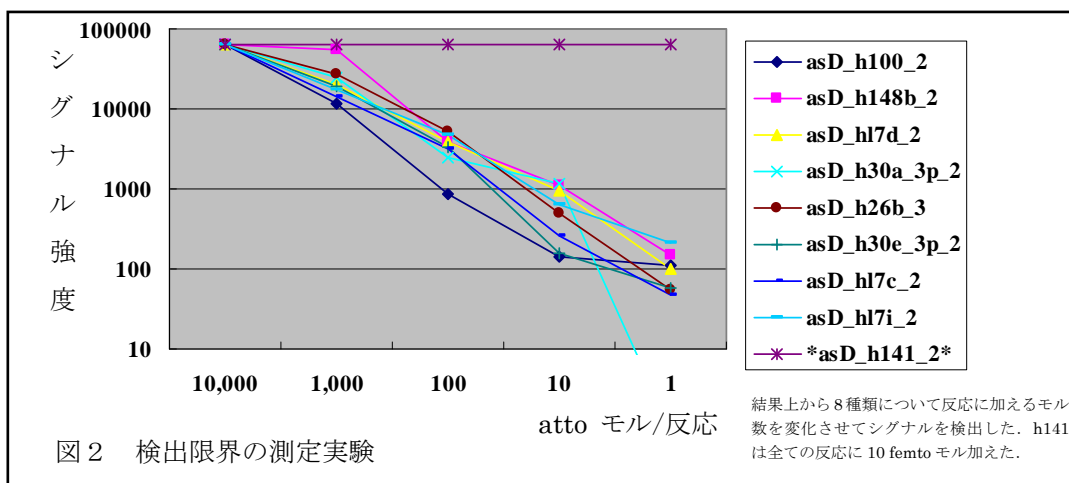
## 2. 2. 2. 2. 2. マイクロアレイ基板を用いた低分子機能性 RNA の高感度検出実験技術の構築

マイクロアレイ基板 S-Bio を用いた MPEX 反応 (DNA チップ研・住友ベークライト共同開発) による低分子機能性 RNA (特に miRNA) の高感度検出技術の開発として、本プロジェクトに於いて4年間研究を進めてきた。本開発では非常に短い RNA 分子を対象としており、その取り扱いを容易にする

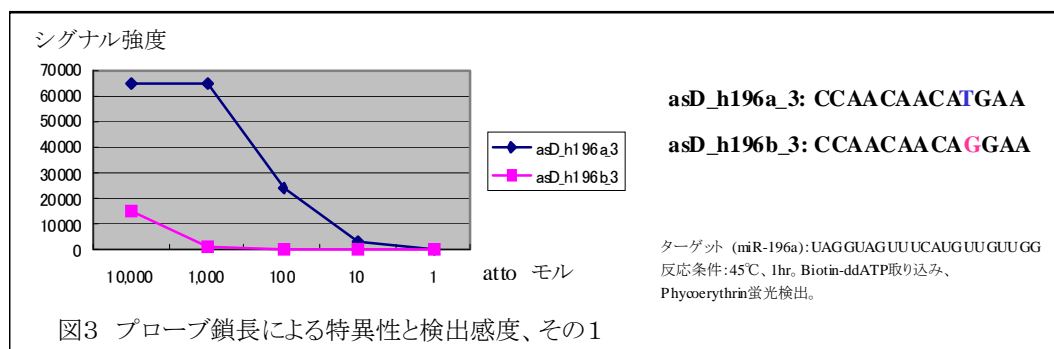


ため、アレイ上にて直接短い RNA を反応させる逆転写酵素を用いた実験系を採用した (図 1)。アレイ表面に固定化されたプライマーを用いて伸張反応を実現することにより、多種類の分子を一度に解析することが可能になる。標識の取り込みはビオチン化 ddATP・ddUTP を用いて miRNA 1 分子あたり 1 分子の標識を行い、プロブの設計に於いてその 3' 末端に A または T を取り込むようにした。長さは特異性を持たせるために後述する項目 2. の実験系により実証しながら調整した。また全ての実験において生化学的な反応の評価をするために、合成 RNA にて miRNA と全く同じ配列を作成して、反応液中のモル数 (分子数) をコントロールしながら解析を行った。

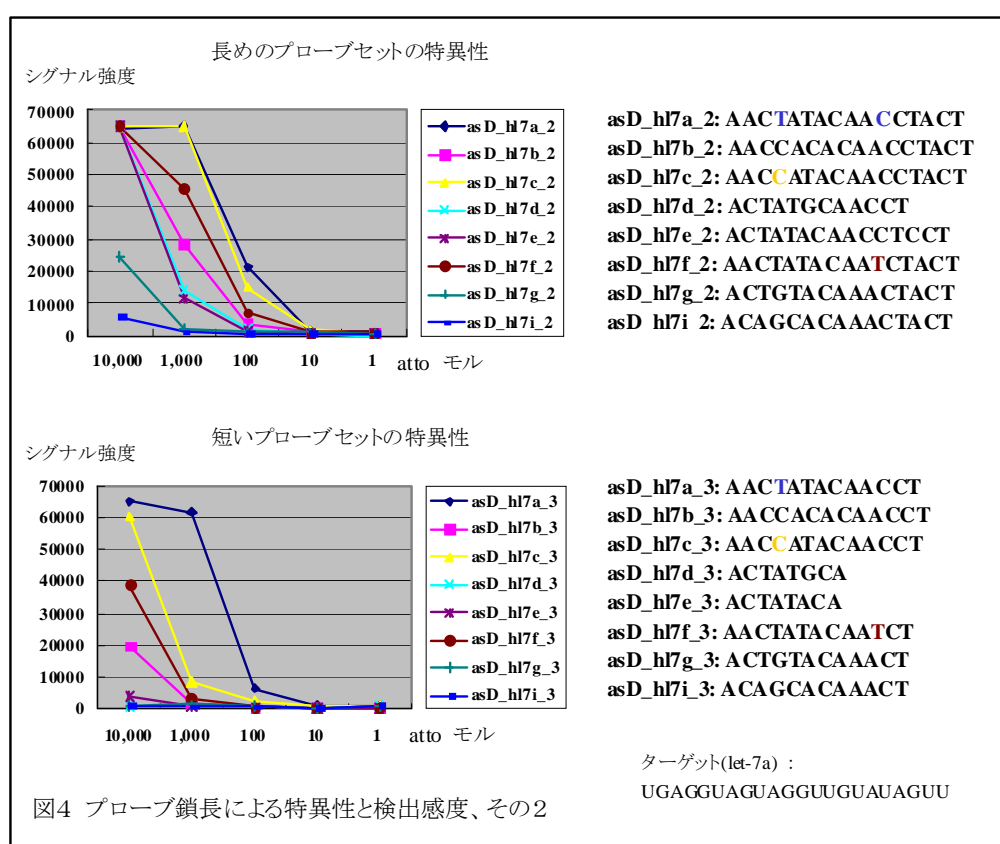
まず目標の項目 1. であるが、図 2 に示すようにほぼ全てのプロブについて 10 atto モルにて、バックグラウンドノイズの 100 付近より有意に高い値になった。1 atto モルで検出できるものは 1/5 ~ 1/10 程度しかないが、それ以外のプロブでも数 atto モルまでは検出可能であり、当初の目標を平成 19 年度の半ばにしてクリアすることができた。h17i (let-7i) や h148b のように 1 atto モル付近でもシグナルが検出できるプロブもある反面、h100 の様に 10 ~ 1 atto モルで数値が変わらなくなってしまうものもあり、プロブの再設計を行う必要がある。19 年度後半からは検出感度の向上を目的として試薬類および反応条件の最適化に取り組み、感度向上とノイズの低減により 1 atto モルで検出可能なプロブの割合が増加しつつあるため、今後も改良を重ねて感度の改善を図る。



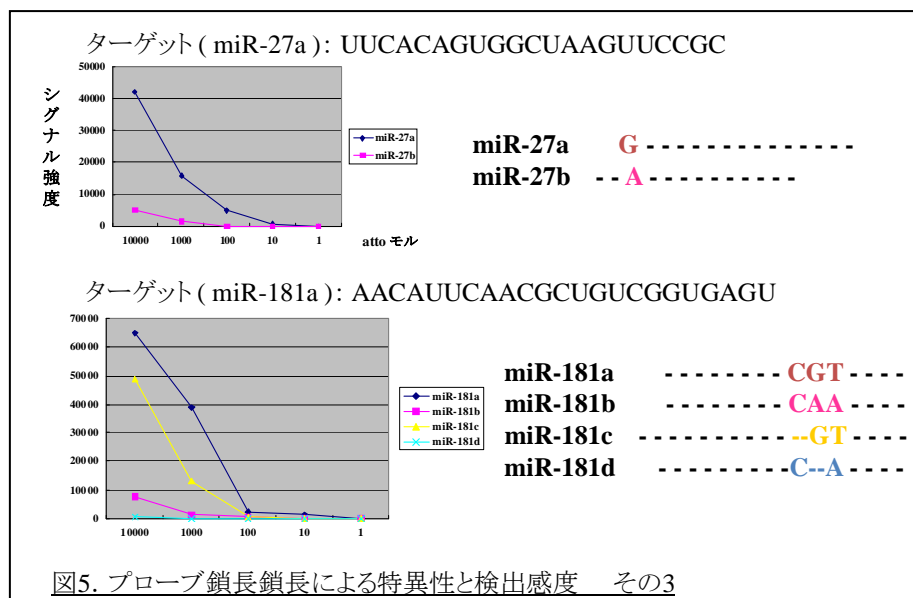
次に項目2.の網羅性と特異性については、特異性を出すためのプローブ設計が可能かどうか重要なため、各 miRNA に関して似たような配列を持つ組み合わせのリストを作成し、各分子の間で非特異的なシグナルが検出されるかどうかを検証した。まず図3のプローブ h196a については、ほぼ中央に1塩基のみ変異のある196b に対して、長いプローブでは特異性が出ないが（データ非表示）、短い13塩基のプローブでは特異性を持って検出できた（図3）。10 atto モルでも蛍光強度にて2000程度のシグナル値が検出されたため、検出限界も数 atto モル以下と問題なかった。



更に複雑な let-7 ファミリーに関しては図4に示す様に、長めのプローブ（図4上）では a と c が区別できないが、短い13塩基のプローブ（図4下）では10倍程度のシグナル値の違いとして判別可能であった。ただし let-7 の場合には検出の特異性を出すためにプローブを短くすると検出限界が10 atto モル程度になってしまい、数 atto モルの感度を保証するためには特異性を多少犠牲にしなければならない。製品化に際しては長短両方のプローブを搭載し、研究対象によってどちらのデータを採用するかを選んでもらうような、解析上の工夫を施す必要がある。



それ以外のプローブに関しても実験を進行中であり例を図5に示した。miR-27では、変異の箇所が5'側になるため、変則的なプローブの配列で対応している。27aと27bは10倍程度の差で特異的に検出が可能であった。一方miR-181では、4種類のファミリー遺伝子の配列中で3塩基の部分に互いに変異がある。181aとcでは1塩基の欠失(挿入)変異があるが、半分弱程度クロスしてしまう結果であった。27も181も、10attoモルで500以上のシグナル値を示しており、検出限界は数attoモルであった。

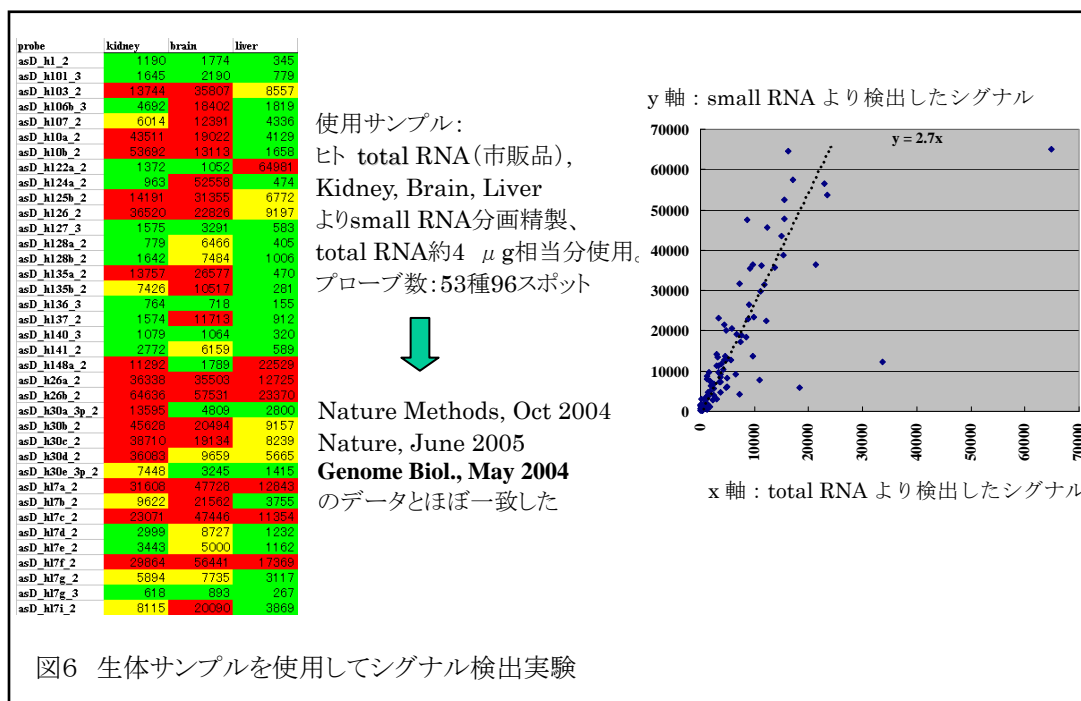


平成 19 年 8 月に引き続き、平成 20 年 4 月と半年ごとに miRBASE (miRNA のデータベース) が大きくバージョンアップを繰り返しており、ヒト miRNA の数もそれまでの 470 種類程度から、900 種類以上へと増加しつつある。そのため mature (major) に分類されたものをほぼ全てと、star (\*) の一部を含む約 800 種類にまでプローブ設計を継続してきた。各分子の間で非特異的なシグナルが検出されるかどうかを検証し、再設計を行うというサイクルを繰り返しながら、使用できるプローブを増やす作業を続けている。またシグナル値の標準化を目指した内部コントロールの検討も行ったが、様々な small RNA について検討を行ったものの最良な分子が特定できなかった。アジレント社の miRNA アレイでも内部コントロールは搭載されておらず、標準化に疑問を呈している現状があるため、引き続き状況を見極めながら対応していく予定である。

## 2. 2. 2. 2. 3. Total RNA と small RNA 分画を実サンプルとして用いた実験結果の検討

最後に項目 3. の実験操作の簡便性であるが、実際に生体サンプルを用いて反応を行なった (図 6)。左図は簡易カラムによる small RNA 分画を直接反応系に入れた場合のシグナル値を表にしたもので、各臓器におけるシグナル強度のパターンは既報論文とほぼ一致した。右図は small RNA と total RNA を反応した場合のシグナル値の相関図であるが、右下の 2 点 (同一プローブ) を除いてほぼ相関していることが分かった (右上の 1 点はシグナル値が飽和していたため本来の値は不明)。相関しないプローブについては、total RNA に含まれる様々な種類の mRNA 分子、あるいはリボゾーム RNA 分子等によりシグナルが検出されていると考えられ、設計を変更して再度検証実験を行っていく必要がある。対象とする miRNA

の種類が増えた際には同様の実験にて検証を行い、必要があれば再設計へと同様のステップを踏んでいくことになる。



また変動が特徴的な 10 種類の miRNA 分子について、TaqMan PCR による検定も行い、アレイ上での MPEX 反応と定量的 PCR 実験という全く異なる実験系において、発現変化について同じ傾向を確認することができた (図 7)。本開発アレイと TaqMan のデータは、発現比の値が TaqMan の方が大きめに出る傾向があるが、ほぼすべてのプローブに於いて相関が見られた。一方現在最も競合する製品と考えられるアジレント社の miRNA 解析アレイツールでは、TaqMan データとの相関はあるものの発現比の値は TaqMan より大小ばらついていた。その結果、本開発アレイとアジレント製品とのデータ間の相関性はあまり見られなかった。



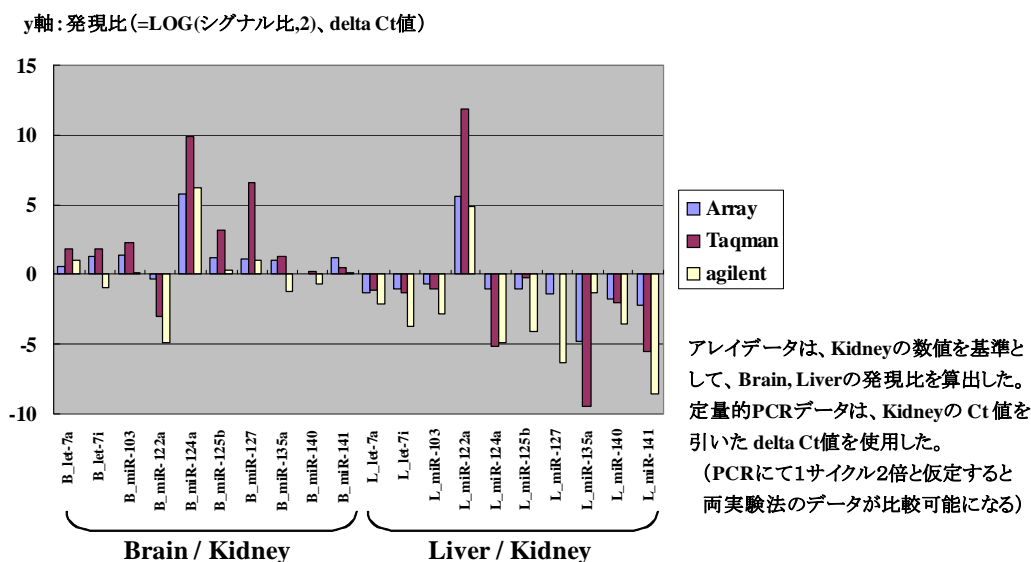


図7. アレイ実験と定量的PCR(Taqman)実験結果の比較

本開発アレイでは total RNA と small RNA でデータが相関していることを図6で示したが、アジレント社では small RNA の使用を推奨していない。図8に total RNA、small RNA の両データのプロット図を示した。特に下段の中～低発現（シグナル）領域に於いて、本開発アレイでの相関性の高さを確認できた。アジレント社のデータから何が起きているのかを推測することは難しいが、本開発製品のデータの明瞭性を示すことができた。

図7に於いてPCRのデータと各アレイ実験のデータは良く相関しているように見える。しかし2種類のアレイデータ間の相関性は必ずしも高くない。その相関を確かめるために図9のようにプロットしてみたところ、PCRで検証したmiRNA分子以外も含めた解析でも相関性はほぼ無いという結果であった。これは、アジレント社のアレイでは total RNA と small RNA で実験結果が大幅に異なる現象が一因ではないかと推測できる。(注：アジレント社では total RNA のみを使用するよう指定している。本実験でも total RNA を使用した。)

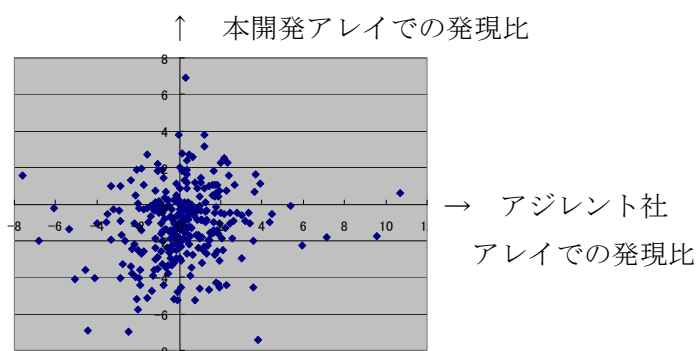
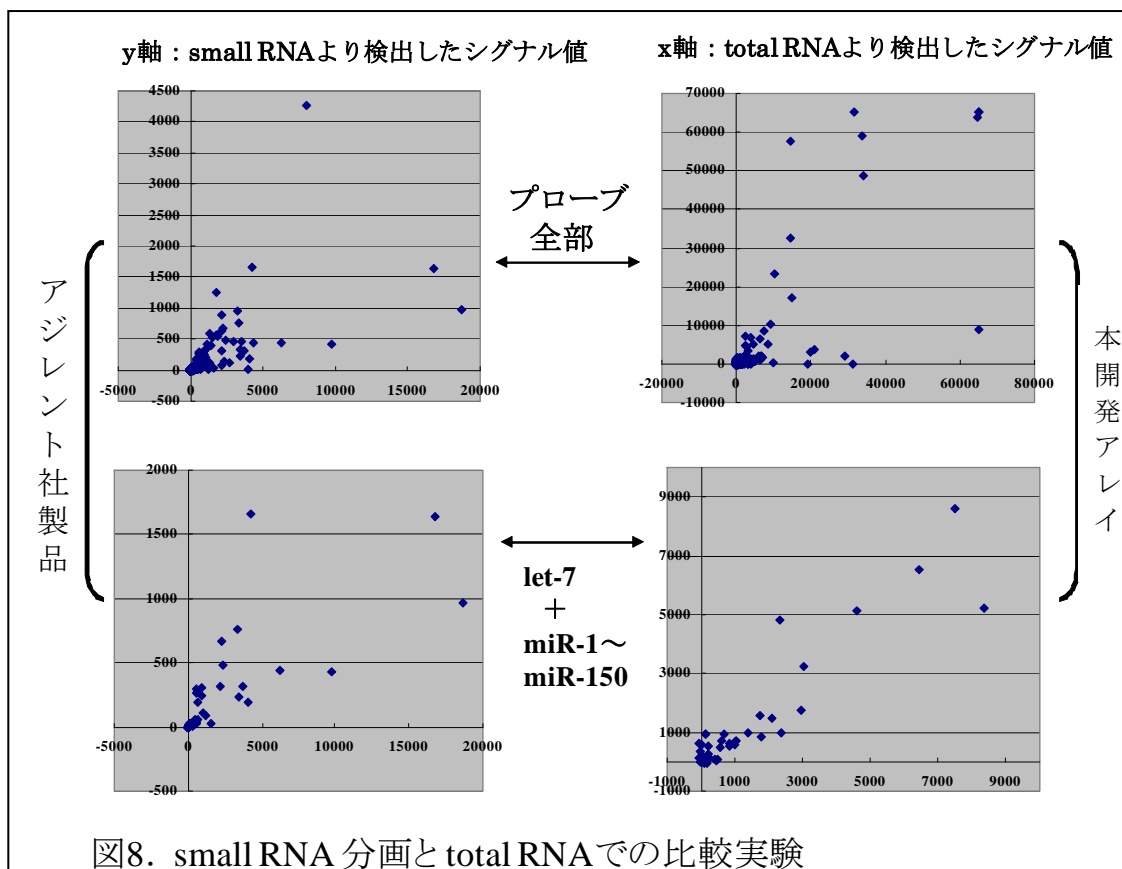


図 9. 2 種類のアレイでの miRNA 発現比の比較解析

アジレント社 miRNA アレイと本研究で開発したアレイを用いて臓器間の発現比 (log2 値) を算出し、それぞれ X 軸、Y 軸にプロットした。

この結果が本開発ツールの優位性を示すものかどうかは議論が分かれることになるが、既存ツールとデータの互換性がほぼ無いということは、すでに miRNA アレイを使用している研究者が、本開発ツールに乗り換えることはほぼ不可能であると想像できる。冒頭にも述べたとおり、サンプルを保有・保管しているが、アレイ実験の経験がない研究者をターゲットに、開発の的を絞ることが肝要であることが明らかになった。

開発項目としては挙げなかったが、miRNA 分子の検出実験を行う上で、実際に機能する

20 塩基程度の mature miRNA の他に、生体内には未成熟の pre-miRNA 分子も存在する。pre-miRNA 分子の配列中には mature miRNA の配列が含まれるため、シグナル検出の際に両方の分子を検出してしまう可能性がある。これを検証するために、ツール開発チームの日本新薬との共同研究として、4 種類の pre-miRNA 分子と同じ配列を長鎖 RNA 合成して供給して頂いた。図 10 のように、mature miRNA 分子と比較して、pre-miRNA 分子は数十分の一～数百分の一以下のシグナル値しか示さず、ほとんど検出されないことが明らかになった。Pre-miRNA 分子はヘアピンループの二次構造を取っていることが知られているため、反応条件によっては二次構造がほどかれてシグナルが検出される可能性があると考え、反応前に熱変性を行ったのちに逆転写反応を行ったり、逆転写反応を酵素の活性限界である 60℃ 前後まで上げて反応したりしたが、いずれも pre-miRNA ではシグナルが検出されなかった。Pre-miRNA 分子を検出するためには、二次構造をほどいた状態で安定化させるような、例えば mature miRNA の配列以外の部分をカバーするカウンターオリゴ DNA などの特殊な処理が必要と考えられるが、まだ検証できていない。以上より、本開発ツールでは、成熟タイプの mature miRNA 分子のみを検出することが明らかになった。

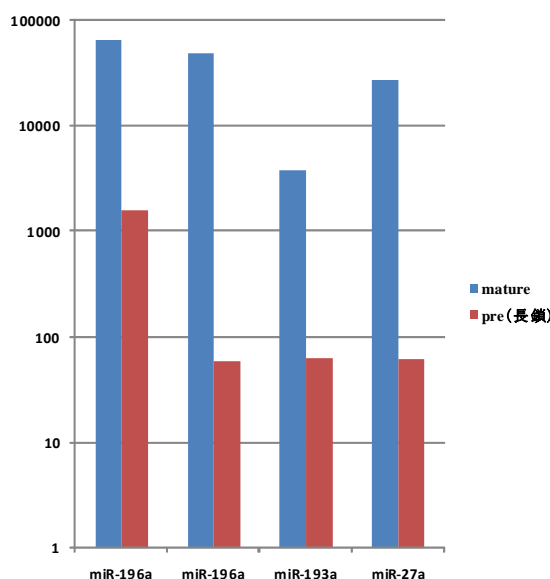


図 10. mature miRNA と pre-miRNA 分子を用いたシグナル検出実験

mature miRNA のシグナルと青いバーで、pre-miRNA (長鎖) のシグナルを赤いバーで示した。縦軸は蛍光のシグナル値で、100 以下はバックグラウンドレベルである。

#### 2. 2. 2. 2. 4. 実サンプルを用いたシグナル検出実験

後半 2 年間は、実サンプルを用いたシグナル検出実験を推進し、miRNA の発現変動を解析することを主目標とした。まずヒトサンプルについては市販の各臓器 total RNA および精製 small RNA 分画、培養細胞から抽出した total RNA を用いて実験を行った (図 11)。

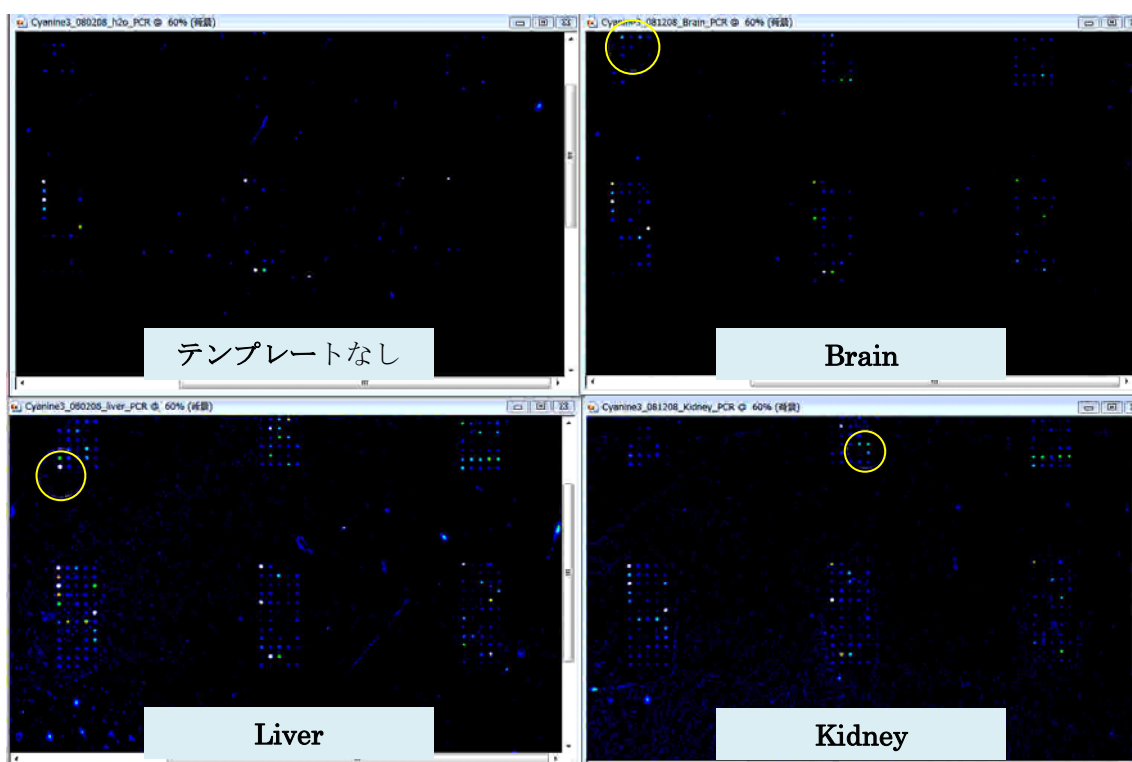


図 11. Brain, Liver, Kidney の total RNA を使用したシグナル検出実験結果

コントロールとしてテンプレートなしの実験も行った。組織ごとに異なるシグナルが幾つか見られ、代表的なスポットを黄色い丸で示した。

更に、ヒトとマウスでは miRNA 分子の約 70% が共通であり、マウスをサンプルとして使用することも想定しているため、ヒトとマウスの臓器での miRNA 検出実験も行った。図 12 はその一例で、肝臓での結果を示した。

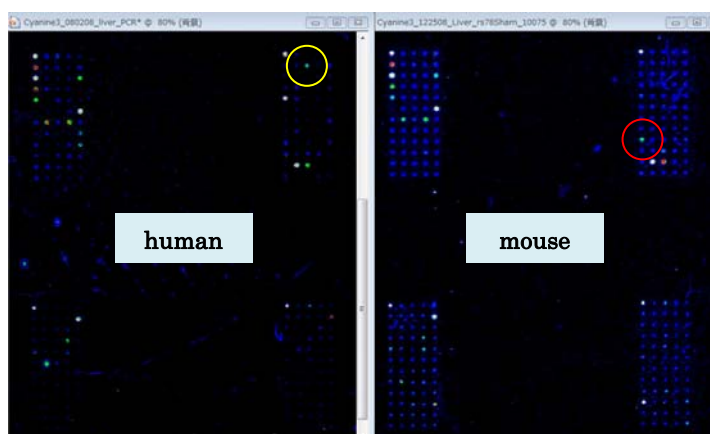


図 12. ヒトおよびマウスの肝臓を用いた miRNA 発現量の解析

ヒト (左側) で発現量が多いものを黄色の丸で、マウス (右側) で発現量の多いものを赤色の丸で示した。

発現している分子のパターンはよく似ているが、その中でもヒトあるいはマウスに特異的な分子が発現しているのが分かった（注：マウスではシグナルの得られないスポットもあるので画像だけでは判別不能である）。

マウスにて miRNA 分子を検出可能であることが明らかになったため、研究代表者が以前に共同研究していた、ストレス前後での遺伝子発現変動解析に用いていたサンプルを使用して、miRNA の発現変動について実験を試みた。血液、脳、肝臓、腎臓、肺について実験を行い、そのうち血液（図 13）、肝臓（図 14）、肺において良好な結果が得られた。残りはサンプルの長期保存により RNA そのものが劣化した可能性もあり、今後の検討課題である。



図 13. ストレス負荷マウスの血液を用いた miRNA 発現変動の解析

ストレス前（上）で発現量が多いものを黄色の丸で、ストレス後（下）で発現量の多いものを赤色の丸で示した。

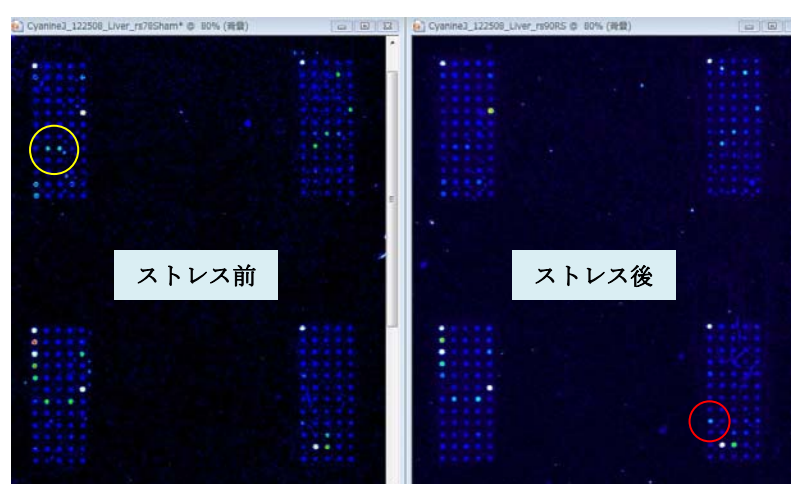


図 14. ストレス負荷マウスの肝臓を用いた miRNA 発現変動の解析

ストレス前（左）で発現量が多いものを黄色の丸で、ストレス後（右）で発現量の多いものを赤色の丸で示した。

以上の実験に用いたストレス負荷マウスのサンプルは既に 5 年以上保存されており、かつメッセンジャーRNA (mRNA) の発現解析に既に使用したものであるためサンプル量に限り

があり、パイロットデータとして解析することは可能であるが、他の実験法での確認実験など更なる解析に発展することは難しい。幾つかでも有意な miRNA の変動が観察された場合には、ストレス関連で新たな共同研究先を探す必要がある。

また、今回用いたサンプルはストレス前後という、コントロールとエンドポイントのみでの実験系である。miRNA は多くの論文で生体器官の発生分化に関わっていることが報告されているため、時系列での発現解析能力が必要になることは明白である。この点に関しては学内に、ヒトおよびマウスの免疫細胞を用いた分化誘導実験系をテーマにしている研究室があり、指導を仰いでいる。図 15 にヒト骨髄系がん細胞である HL-60 を用いた分化誘導に伴う miRNA 遺伝子発現解析実験を示した。

本開発研究中ではまだ予備的な実験までしか到達できなかったが、培養および分化誘導実験に必要な器材は学内にて調達できる目処が立ったため、分化誘導に関わる miRNA 分子の発現変動について時系列での解析に取りかかる予定である。

またデータは示さないが、微量サンプルにも対応するため、血液一滴（数十  $\mu$ l 程度）より small RNA 分画を精製して、miRNA の発現量を測定できた。最近の報告で各種の体液中に miRNA 分子が分泌されているとの記述があるため、その役割を究明するためのツールとしても使用できることが明らかとなった。

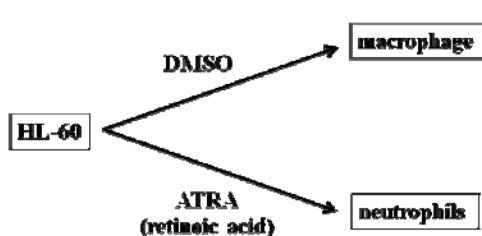
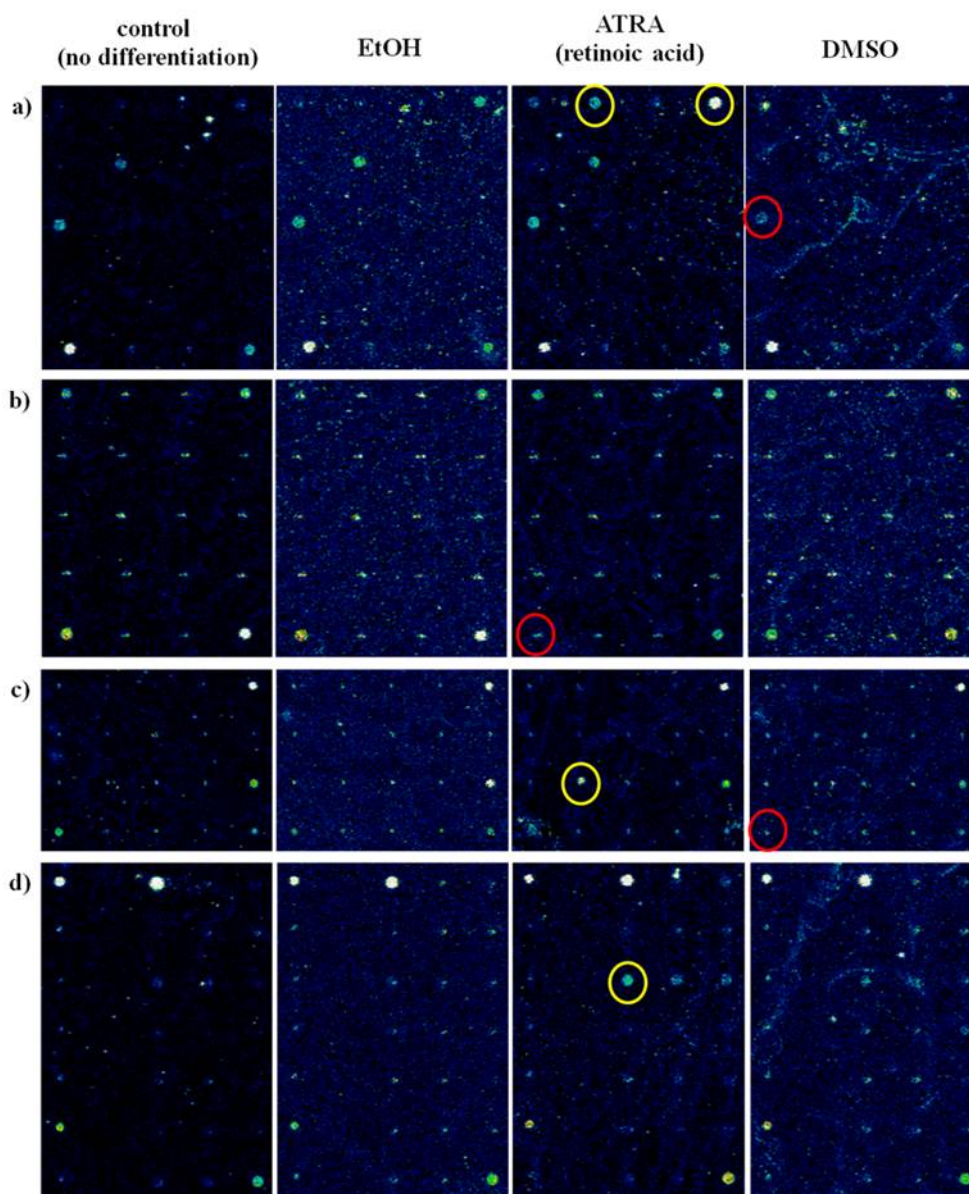


図 15. ヒト HL-60 細胞の分化誘導に関わる miRNA の解析  
 ATRA (レチノイン酸) により好中球に、DMSO によりマクロファージに分化する。EtOH は ATRA の溶媒。特異的に発現が上昇した miRNA のスポットを黄色い○で示し、発現が下降した (抑制された) miRNA を赤い○で示した。カラム a~d はアレイ中の別のスポット位置を表す。

## 2. 2. 2. 2. 5. 検出結果の可視化についての検討

上にも述べた通り、既存の miRNA 検出ツールが競争相手として存在するため、項目 3. の特徴を更に生かして一般の研究者に広く受け入れられるツールを開発する必要がある。平成 19 年度末までに、蛍光検出の手順をストレプトアビジン-IP と BCIP-CIP による色素沈着反応に置き換えることにより、ターゲット濃度に依存して色素が沈着することを確認できた。その後、検出感度向上や検出のダイナミックレンジを拡大するために様々な実験条件の検討を行ったが、若干の向上が見られるにとどまった。しかし沈着色素を画像データとして取り込む方法として、PC 用スキャナーを用いて画像を取り込むことが可能になり、画像の解像度および濃淡の階調が数段上がった。顕微鏡のデジタル取り込み画像（画像調整済み）と、スキャナーでの取り込み画像を図 16 に示す。顕微鏡では解像度が低く（性能に依存）、また画像の四隅での画像のゆがみも見受けられたが、スキャナーでは平面的に画像をスキャンするため問題なかった。図 17 には検出感度のグラフを示した。

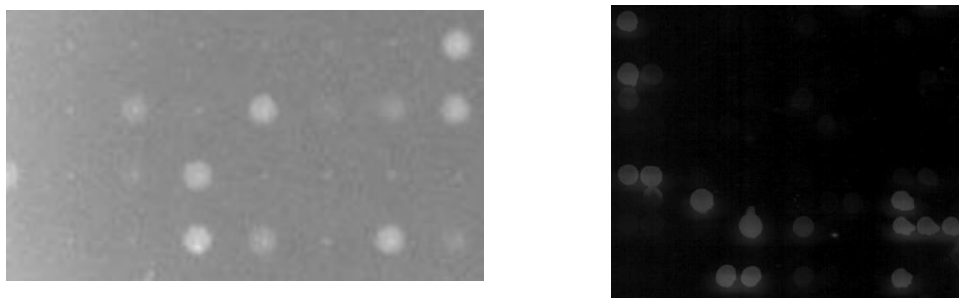


図 16. 色素沈着実験例の顕微鏡画像（左）と PC 用スキャナー画像（右）の比較  
両方ともネガ画像に変換しているが、顕微鏡画像のみコントラスト調整を行っている。

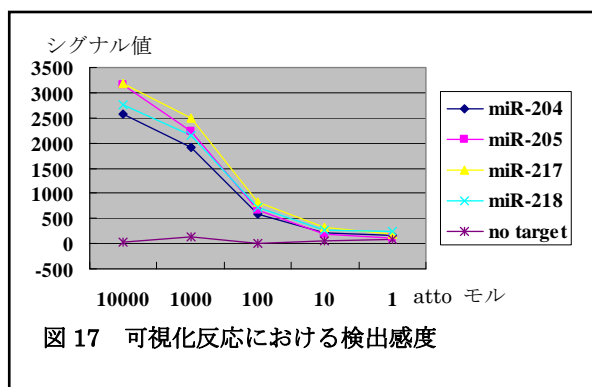


図 17 可視化反応における検出感度

ただし、PC スキャナーの場合にはスライドガラス本体の反射光や、本体が透明なため露光調整がうまくいかないことがあり、ソフトウェア上での設定が必要になる。この点を解消する目的もあり、また沈着色素の量を画像変換する際の効率も考えて、現在住友ベークライトで開発中の白色スライドガラスの使用を検討している。現段階では耐熱温度が 80℃ のため、逆転写反応後の完全洗浄ステップでの加温が十分にできないが、プラスチックの配合調整により耐熱温度を上げられるよう改良を求めている。



## 2. 2. 2. 2. 6. 反応のサイクル化による感度向上に向けた技術開発

現在の逆転写反応は1回のみであるため、通常のハイブリダイゼーションによる検出よりは高感度であるが、ABI社のTaqMan miRNA PCRキットには到底及ばない。現在は開発目標のattoモル(10の6乗分子)レベルでの検出感度を何とか達成できたところである。この逆転写反応をサイクル化することができれば、PCRのような指数関数的ではないが、リニアにシグナルを増強することが可能になる。それが理論的に可能になるのは、現在のプローブ(プライマー)設計が $T_m$ 値を40-45°Cに設定しており、使用している逆転写酵素の耐熱温度が60-65°Cであることによる。つまり、60°Cと42°Cの2ステップのサイクルを完全に温度調節できれば、熱変性とアニール・酵素反応は可能になると考えられる。

その実現には技術的な開発が必要になる。マイクロアレイは平面上での反応になるため、反応液の温度を完全に制御することが非常に難しい。住友ベークライトの試作機で、アレイの両面からヒートブロックを押し当て、各ヒートブロックの間を移動していく器械により実現に一歩近づいた。しかし実際に反応を行うと、数十マイクロリットルの反応溶液の中でかなり温度にムラがあることが分かった。現在はスライドガラスそのものにも加工を施し、また反応溶液を密封するシール素材の検討も行い、ようやく反応溶液の温度コントロールが可能になってきたところである。ただしまだ量産できる体制ではないため、今後もし引き続き反応条件などを検討して、検出感度の向上に努めていく。

## 2. 2. 2. 2. 7. まとめ

### (1) まとめ

これまでの研究開発によって、当初の3つの目標項目は全て達成できた。また精製total RNA、small RNA分画を用いたシグナル検出実験においても十分に発現解析が可能であることが示された。また現在進行中である培養細胞を用いた分化誘導実験系に於いて、分化に関わるmiRNA遺伝子群の抽出を試みている。以上の実験開発データを学会発表しており、今後論文などで報告することにより本ツールの有用性を認識してもらえよう努力し、今後の製品化に踏み出せるよう研究を継続する。現状での1サンプルあたりの実費は1万円前後であり、製品化に際しても他のキット(クローニングキット等)と同程度のコストでマイクロアレイ解析が可能になると想定される。現在進行中の応用実験例の提示と、可視化技術の実用性が示されれば、本研究開発の最終目標である、「マイクロアレイ実験に精通していない研究者が、通常の酵素反応等と同様に一般的な実験法として採用し、手持ちのサンプルでmiRNAの網羅的発現解析を行い、有用なmiRNA分子を抽出できる」安価な国産技術のツールとして市販化でき、かつ普及する可能性が大きいと考えている。

### (2) 実施計画書記載の達成目標に対する到達度

目標であるmicroRNA検出のための基礎技術の開発項目として、1. 検出感度、数attoモルレベル、2. 網羅性と特異性のバランス追求、3. 簡便な実験系の開発、の全ての項目に関してモデル実験系を用いて目標値を達成できた。今後の製品開発に向けて生体サンプルでの実験系も行い、基礎段階をクリアできたと考えている。

研究応用の一例として培養細胞の分化誘導系を用いて、時系列での miRNA 発現解析研究を遂行中であり、ツールとしての性能評価についても今後推進していく必要がある。最後に、既製品との違いを明確にするためにも、項目 3. に関して簡便性を追求するためにレーザーキャナ等のマイクロアレイ専用機器類を使用しない実験検出系を検討することが求められる。以上により、日本独自のツール開発を目指して今後も努力して行きたい。

## 2. 2. 3 RNA の新規合成基盤技術開発と化学分子設計

集中研②、分室 7（日本新薬）、分室 8（ヤマサ醤油）

共同実施先： 産総研②、東京大学（4；和田）

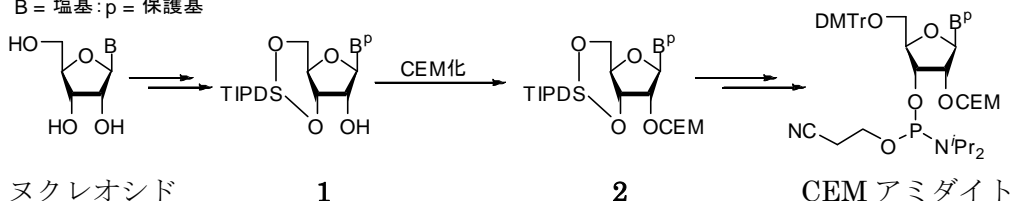
近年 RNA の生命現象に対する役割が次々と明らかにされている。これに伴い、生体内での役割を解析するツールとして純度の高い化学合成 RNA の必要性が高まっている。我々は RNA 合成法を改良し 2-cyanoethoxymethyl (CEM) 基を新規保護基とした新規 RNA 合成法 (CEM 法) を開発したことを既に報告している<sup>1)</sup>。CEM 基の特長は従来の RNA 保護基と比べ、立体障害が非常に少ない点にある。これは化学合成において 1 塩基ずつアミダイトをつなげていく際の収率へと反映される。また、立体障害が小さいことを利用して、従来困難であった修飾型核酸を化学合成することも可能となる。我々は、この新規核酸合成技術を土台として、新たに見出された機能性 RNA の生体内での機能を検証するとともに、CEM 基の反応性の高さを活用した修飾型核酸を創製し、将来の機能性 RNA の医薬品化に応用できる基盤技術を開発することを目的として検討を行った。プロジェクト前半では、原料アミダイトの合成法、および、長鎖機能性 RNA 合成条件の検討を行い、プロジェクト後半では、修飾体や長鎖合成 RNA の応用を中心とした新規機能性核酸の合成検討を行い、所定の成果を得ることが出来た。本技術は、プロジェクトの成果を、抗ガン剤、抗ウイルス剤といった核酸医薬品や再生医療に応用し産業化するために不可欠の基盤技術である。

### 2. 2. 3. 1. 原料となる CEM アミダイト合成法の確立（集中研、日本新薬）

新規合成法により RNA を大量に化学合成するためには、その原料となるモノマーブロックすなわち 2-cyanoethoxymethyl (CEM) アミダイトを大量に調製する必要がある。そこで、これを合成する上で鍵反応となる CEM 基を導入する反応（1 から 2 への反応、以下、CEM 化反応と記載）を検討した。この CEM 化反応以外の工程は従来の方法論と類似の条件を適用できるため、この工程を改良することができれば、スケールアップも問題なく対応できる（スキーム 1）。

#### スキーム 1

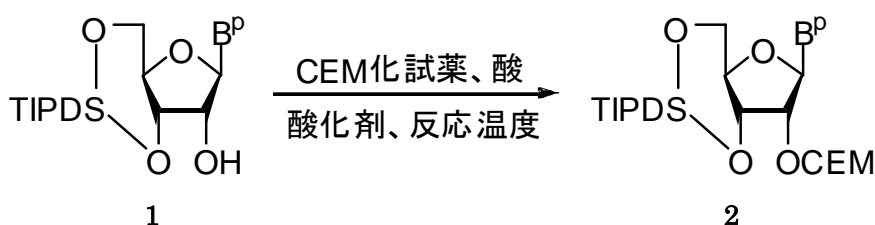
B = 塩基; p = 保護基



既に報告した CEM アミダイト合成法<sup>1)</sup> (*Organic Letters*, 2005) における、CEM 化反応

条件では、次のような問題点があった。1) 反応に用いる試薬が高価であり、発煙する程度活性が高く、扱いには特に注意が必要である。2) 反応を超低温で制御しなくてはならないため、スケールアップが困難である。3) 塩基の種類によっては収率が低いものがあり、異なる合成ルートでの合成が必要であった。そこでこれらの問題点を解決するために、1) 反応性が高く高価な試薬類を使用しない方法に変更すること。2) スケールアップが容易とするために反応制御の容易な温度で効率良く反応が進行すること。および、3) 塩基の種類によらず同じ合成ルートで合成できること。以上3点を目標として条件検討を行った。その結果、反応試薬としては、安価で反応性が低く取り扱いが容易なメタンスルホン酸とヨウ素を使用できることを見出した。また、反応温度も従来法より制御が容易な条件である摂氏0度で、塩基の種類によらず高収率で目的とする化合物2を得ることに成功した(スキーム2)。この結果は、平成18年8月に特許出願を行った(特願2006-210439号)。

スキーム2



	酸	酸化剤	反応温度	収率
旧条件	トリフルオロメタンスルホン酸 <sup>※1</sup>	N-ヨードスクシニミド <sup>※2</sup>	-40℃	45-89%
新条件	メタンスルホン酸 <sup>※3</sup>	ヨウ素 <sup>※4</sup>	0℃	80-88%

※1 非常に強力な酸で発煙を生じ、通常はアンプルにて保存。日本では劇物に指定。

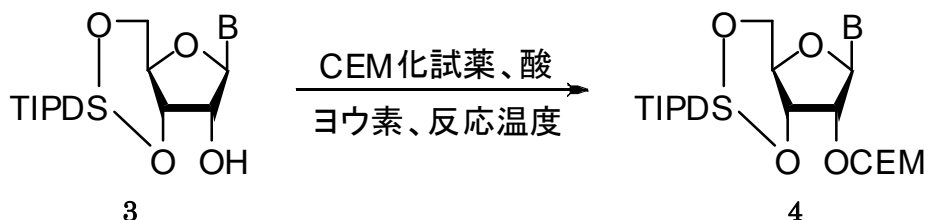
※2 反応性が高く実験室レベルで広く使用されている。反応性が高い反面、試薬の安定性が低く冷凍保存する必要がある。非常に高価である。

※3 発煙は生じず、アンプル保存の必要はない。劇物に指定されていない。

※4 反応性が低く非常に安価である。常温にて保存可能。

次に、CEM アミダイト合成のさらなる改良を目指し、合成ルートの検討を行った。従来の方法では塩基部を保護した後に CEM 基を導入していたが、CEM 基をまず導入してから、塩基部を保護するルートを検討した。この利点として次のことが挙げられる。1) 合成中間体(3, 4等)の結晶性向上により、精製が簡便化できる。2) 塩基部保護基の脱離を回避できる。3) 塩基部保護基導入が簡便化できる。4) 塩基部無保護のモノマーブロックの合成が可能となる。塩基部無保護による CEM 化反応を検討した結果を以下に示す(スキーム3)。

スキーム 3



Entry	塩基	反応温度	酸	収率
1	シチジン	0 °C	トリフルオロメタンスルホン酸 (触媒量)	88%
2	アデノシン	-40 °C → -10 °C	トリフルオロメタンスルホン酸 + BF <sub>3</sub> · Et <sub>2</sub> O	94%
3	グアノシン	-40 °C → -15 °C	トリフルオロメタンスルホン酸	41%

シチジン誘導体については、従来に比べ穏和な条件で同等の収率で目的物を得ることができた。アデノシン誘導体においても塩基部が反応しないように条件をコントロールすれば、従来と同等の収率で目的物を得ることができた。本反応ルートは上述のルートに比べて効率的であり、有用性が高いと考えられる。この結果は、平成 19 年 1 月に特許出願を行った(特願 2007-011813 号)。

最後に、原料合成法が確立したことから、RNA 合成の「グラムスケール」までのスケールアップ検討を行った。一般に、「実験室レベル」から「グラムスケール」までスケールアップ出来れば、さらにスケールをあげて、「工業スケール」で合成をおこなうことは、ほぼ障害無く進めることが可能である。CEM 法を用いてグラムスケールで RNA を合成した場合、従来法 (TBDMS 法) と比べ約 3 倍以上の収量が得られることを確認した (データ示さず)。

以上、プロジェクト前半では天然型 RNA 合成のための原料アミダイトの合成法確立に注力し、工業化可能である合成法をほぼ確立することが出来た。

### 2. 2. 3. 2. 長鎖機能性 RNA (110-mer pre-miRNA) の化学合成 (集中研、日本新薬)

CEM 保護基の特長は従来の RNA 保護基と比べ、立体障害が非常に少ないという点にある。これは化学合成においてアミダイトをつなげていく際の収率向上に反映されるため、数多く連結反応を繰り返す長鎖の RNA 合成において CEM 法は従来法より圧倒的に有利である。従来の RNA 合成法では、鎖長が 30 程度までの短鎖 RNA 合成は可能であるが、鎖長が 50 以上の長鎖 RNA オリゴマーの合成は非常に困難であった。今回、CEM 法の特長を生かし、これまで化学合成としては世界で報告例がない鎖長が 110 におよぶ長鎖機能性 RNA オリゴマーの化学合成を行った。合成した RNA は機能が確認されている<sup>2)</sup> miR-196a の前駆体 RNA<sup>3)</sup> (pre-miRNA, 図 1) である。



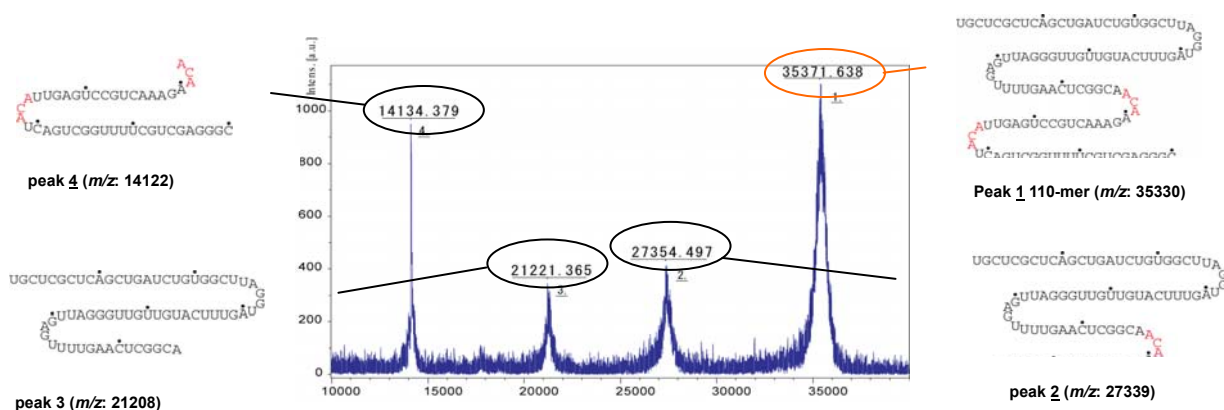


図4. 110-mer pre-miRNA 酵素消化物の TOF-Mass 解析

次に、miR-196a の標的配列を組み込んだルシフェラーゼ遺伝子を用いたレポーターアッセイ系を構築し、化学合成した 110-mer pre-miRNA の生物活性を調べた。図 5A に示すように、110-mer pre-miRNA は標的とする遺伝子の発現を成熟型 22-mer 二本鎖 RNA と同程度に抑制した。この結果から我々は、化学合成した 110-mer pre-miRNA が細胞内で正常にプロセスされ、生じた miRNA が発現を抑制したものと結論付けた。

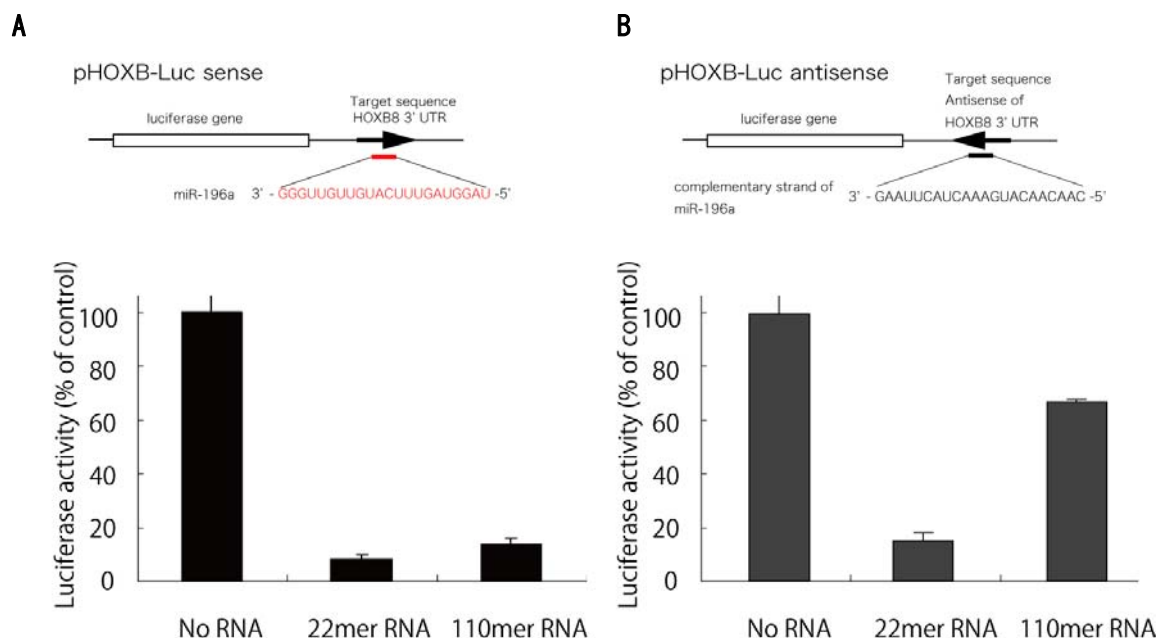


図5. 化学合成した 110-mer pre-miRNA の標的配列のセンス配列(A)とアンチセンス配列(B)をそれぞれ標的として組み込んだレポーター系での発現抑制効果

さらに 110-mer pre-miRNA の標的とする配列の鎖選択性を調べた。成熟型 22-mer 二本鎖 RNA は標的のセンス鎖配列、アンチセンス鎖配列に選択性を示さないのに対し、110-mer pre-miRNA はセンス鎖配列選択的に作用し、発現を抑制することを見出した(図 5B)。この結果は pre-miRNA 特異的な性質を示しており、siRNA を利用した遺伝子発現抑制においてしばしば問題となる相補鎖由来のオフターゲット効果を回避できる可能性を示すものである。

以上まとめると、我々は、CEM 基を用いた新規 RNA 合成法 (CEM 法) を用いることで 110-mer pre-miRNA を高収率、高純度で化学合成した。この配列はこの時点で、世界最長の化学合成 RNA である。長鎖 RNA の構造を確認することは困難であるが、化学合成した 110-mer pre-miRNA を酵素分解と MALDI-TOF MS 分析により構造確認を行った。さらに生物活性を測定し、成熟型 22-mer 二本鎖 RNA とは異なる pre-miRNA に特異的な標的鎖選択性を示すことを見出した。これらの結果から、我々が開発した CEM 法は鎖長が 100 を超える RNA 合成を可能とする実用性の高い方法であり、機能性 RNA 分子を化学合成し、生理機能を解析する方法論が有効である一例を示すことができた。この成果は学術論文として報告した

(*Nucleic Acids Res.*, 2007)。また、長鎖合成 RNA は、共同研究としてプロジェクト内に供給を行い、機能性 RNA の機能解析等に使用した。

### 2. 2. 3. 3. 人工 mRNA の合成 (集中研、日本新薬)

生体内に存在する RNA には 5' 末端に Cap 構造を持つものが存在し、代表的なものが mRNA である。さらに、タンパク質をコードしない機能性 RNA にも Cap 構造を有しているものがあり、Cap 構造を有する RNA の合成は、機能性 RNA の解析において有意義である。また、mRNA の人工的な合成により新たな応用分野を開く可能性もある。我々は、化学合成による長鎖 RNA の利用の一例として、人工 mRNA の合成を行った。合成したのはペプチドホルモンである GLP-1 の配列をコードした mRNA である。

RNA の 5' 末端を Cap 化するためには、RNA の 5' 末端をジリン酸化する必要がある。しかし、ジリン酸化体の作製は、酵素法では適切な方法が無く、化学的にも、2004 年に報告のあった従来法<sup>5)</sup>では副産物が多く、収率が低いためジリン酸化体を得ることは困難であった。そこで、我々は効率的なジリン酸化反応を可能とする新規リン酸化試薬の開発を行った(図 6)。生体内ではリン酸化は免疫反応やシグナル伝達など非常に重要な役割を担っており、開発したリン酸化試薬は合成物質の応用範囲を広げるものと考えている。また、このリン酸化試薬は RNA のみならず、低分子やペプチドなどのリン酸化にも応用可能であり有用性が高いことから、リン酸化試薬およびリン酸化試薬の効率的な精製法について、それぞれ特許出願(特願 2010-10260、特願 2009-279488)を行った。



### 新規リン酸化試薬による モデル配列のリン酸化反応

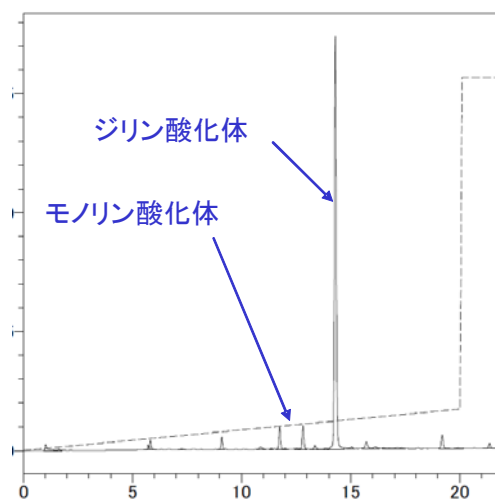


図6. 従来法および新規リン酸化試薬を用いたモデル配列 (3-mer) ジリン酸化反応産物の HPLC(陰イオン交換カラム)による分析

次に、人工 mRNA の合成を行った。作成法の概略を図7に示した。主鎖となる 119-mer～130-mer の RNA を CEM 法で化学的に合成し、次に化学的にモノリン酸化、続いてジリン酸化を行い、ジリン酸化体とした。化学合成したジリン酸化 130-mer RNA は精製後、HPLC で分析し、その純度が極めて高いことを確認した (図8A)。その後、酵素反応にて 5' 末端を Cap 化し、poly(A) polymerase により 3' 末端に poly(A) テール付加反応を行い、5' -UTR の長さや Kozak 配列<sup>3)</sup>の有無を変えた 5 種類の人工 mRNA を合成した。Cap 付加の確認は、5' 末端を含む MazF 切断断片の分子量を MALDI-TOF-MS で測定することで、Poly(A) テールの付加は鎖長をポリアクリルアミド電気泳動で確認することにより行い、それぞれ正しく付加されていることを確認した (データ示さず)。さらに、GLP-1 をコードする 130-mer RNA に CEM 法で 40 mer の poly(A) を付加した 170-mer RNA のジリン酸化体を化学合成し (図8B)、さらに酵素反応で Cap 付加を行い人工 mRNA を作成した。ここで化学合成を行った 170-mer RNA の配列は現時点で世界最長である。

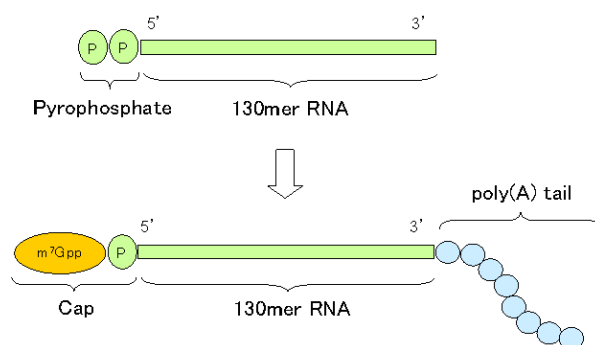


図7. 人工 mRNA の作成

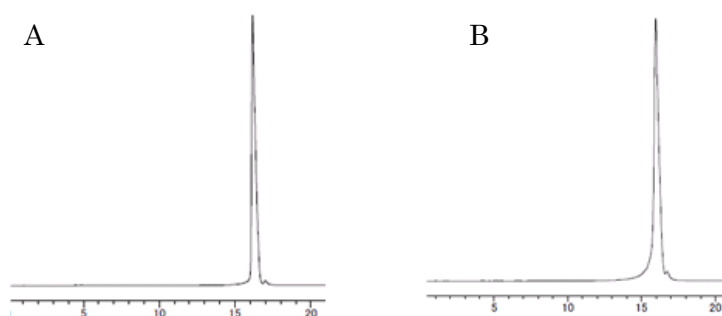


図8. 130-mer (A)、170-mer (B) のジリン酸化体の HPLC (陰イオン交換カラム) による分析

さらに、合成した人工 mRNA を用い、小麦胚芽無細胞発現系で GLP-1 (7-37) の発現を行った。その結果、主に 5 種類中 4 種類で mRNA 濃度依存的な GLP-1 (7-37) の発現を確認することができた (図 9)。さらに、作成した mRNA をエレクトロポレーション法により HEK293T 細胞に導入したところ、培養細胞においても GLP-1 (7-37) が発現することを確認できた (データ示さず)。

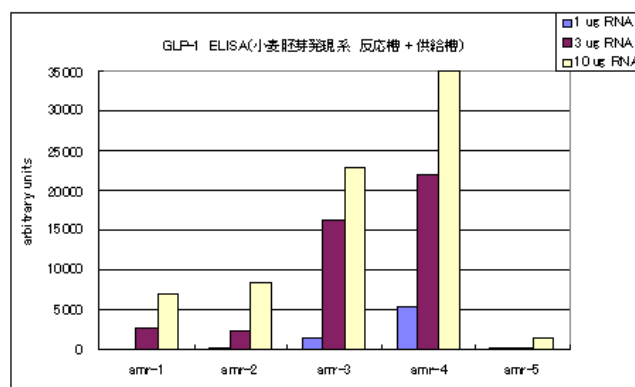


図 9. 無細胞発現系での GLP-1 ペプチドの発現

化学合成により poly(A) を付加した 170-mer mRNA についても小麦胚芽無細胞発現系において GLP-1(7-37) の発現を確認することが出来た (図 10)。

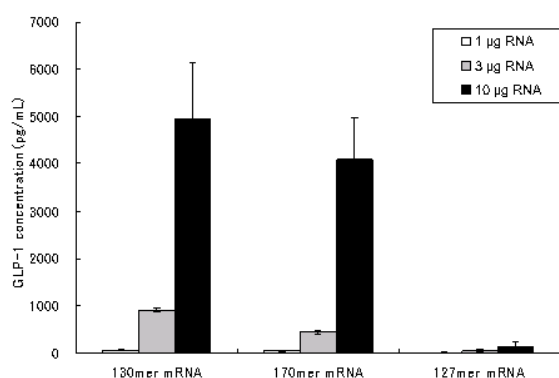


図 10. Poly(A) まで含めた配列を化学合成した 170-mer mRNA の GLP-1 ペプチド発現

以上をまとめると、CEM 法によるこれまでの最長 (110-mer) を超える 119-mer~170-mer の GLP-1 をコードする長鎖 RNA のジリン酸化体を合成した。この過程で、高収率でジリン酸化体が得られる新規のジリン酸化試薬およびリン酸化試薬の効率的な精製法を見出した。次いで、酵素による長鎖 RNA の 5' 末端の Cap 化反応、及び 3' 末端の poly(A) 付加反応を行い、人工 mRNA を合成することに成功した。酵素反応の確認は MALDI-TOF-MS および電気泳動で行い、Cap 構造、poly(A) とともに付加していることを確認した。このようにして作成した mRNA を用いて小麦胚芽無細胞発現系で GLP-1(7-37) の発現を行ったところ、mRNA 濃度依存的な GLP-1(7-37) の発現が認められた。また、作成した mRNA を電圧ポレーション法により HEK293T 細胞に導入したところ、細胞においても GLP-1(7-37) の濃度上昇が認め

られた。

タンパク質をコードしない機能性 RNA と考えられていた RNA が短いペプチドをコードしていた例も見つかっており、また、ショウジョウバエにおける *oskar* RNA のように卵形成の初期には non-coding RNA として働き、胎児形成期には mRNA として働く例も見出されてきていること<sup>6)</sup>から、広義の機能性 RNA 解析のためには Cap 構造を保持する RNA で短いペプチドをコードする可能性をもつ RNA も含めた検討が不可欠であると考えられる。また、人工 mRNA の合成はインフルエンザウイルスや未知の RNA ウイルスによるパンデミックに対応できる RNA ワクチン開発の基盤技術となり、新たな医療への応用につながることを期待される。この成果は学術論文として報告する予定である。

### 2. 2. 3. 4. 非天然型修飾 RNA の合成研究 (東京大学、集中研)

#### (1) ホスホロチオエート RNA の立体選択的合成と性質

核酸医薬品の実用化には、生理活性を有する核酸修飾体の開発が必要である。核酸のリン原子部分は核酸修飾の対象の一つであるが、リン原子にキラリティーがあるため、従来の化学合成法で得られる誘導体は多くの立体異性体の混合物となっていた。リン原子の立体配置によって核酸の熱的安定性やヌクレアーゼ耐性などが大きく異なることが知られているため、それらの立体選択的合成は極めて重要な課題である。我々が開発したリン原子の絶対立体配置が完全に制御されたホスホロチオエート DNA の実用的な合成法であるオキサザホスホリジン法<sup>4)</sup>を応用して、オキサザホスホリジン法によるホスホロチオエート RNA の立体選択的合成法の開発を行なった (図 11)。

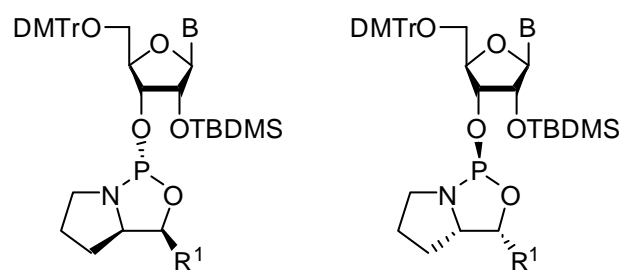


図 1 1. 光学的に純粋なオキサザホスホリジンモノマーの構造

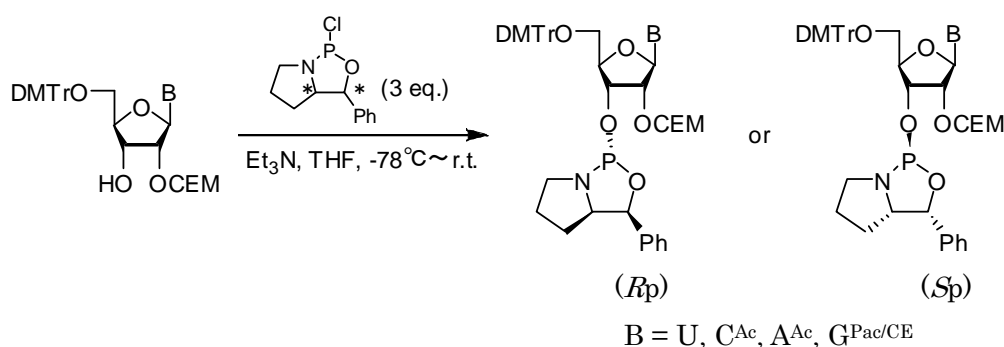
2' -*O*-TBDMS 保護リボヌクレオシド 3' -*O*-オキサザホスホリジンモノマーの立体選択的合成を検討し、4 種類の核酸塩基を有する、光学的に純粋なモノマー *S<sub>p</sub>* 体および *R<sub>p</sub>* 体を良好な収率で得ることができた。次に、液相法および固相法により、ジリボヌクレオシドホスホロチオエートの立体選択的合成を検討し、ApsU、CpsU、UpsU、GpsU それぞれについて、2 種類のジアステレオマーを 99:1 以上の立体化学的純度で得ることに成功した。さらに、10 量体の固相合成を行ない、光学的に純粋な all-*R<sub>p</sub>*-(Ups)<sub>9</sub>U および all-*S<sub>p</sub>*-(Ups)<sub>9</sub>U を得ることができた。2 量体以上の立体化学的に純粋なホスホロチオエート RNA の化学合成は

これが世界で初めての例である。得られたオリゴマーの二本鎖形成能を  $T_m$  測定（高塩濃度条件下）を行なうことにより評価した。all- $R_p$ -(Ups) $_9$ U/(Aps) $_9$ A の  $T_m$  値は 28.9°C であり、天然型 RNA 二本鎖 (Ups) $_9$ U/(Aps) $_9$ A (25.7°C) よりも高い  $T_m$  値を示した。一方、all- $S_p$ -(Ups) $_9$ U は相補的な (Aps) $_9$ A と二本鎖を形成せず、リン原子の立体が制御されていない random-(Ups) $_9$ U/(Aps) $_9$ A の  $T_m$  値は 10.3°C であった。以上の結果は、リン原子の絶対立体配置がホスホロチオエート RNA の二本鎖形成能に大きな影響を及ぼすことを示しており、化学修飾型機能性 RNA の分子設計に対してきわめて重要な知見が得られたといえる。

## (2) 2'-*O*-CEM 基を有するオキサザホスホリジン型モノマーの合成

オキサザホスホリジン法によるホスホロチオエート RNA の立体選択的合成は、これまで一般的な保護基である 2'-*O*-TBDMS 基を用いていた。ホスホロチオエート RNA の立体選択的合成の反応効率を高め、種々の配列ばかりでなく長鎖への応用を可能とするため、RNA 合成において高い縮合収率が期待できる 2'-*O*-CEM 基を有するモノマー  $S_p$  体および  $R_p$  体の合成を検討した。原料に対してキラル補助基を導入したところ、いずれの塩基の  $S_p$  体および  $R_p$  体の合成においても、高立体選択的に (>99:1) 反応が進行することが確認された（スキーム 4）。本モノマーを用いて、UpsU の固相合成を行い、( $R_p$ )-モノマーからは ( $S_p$ )-UpsU が、また ( $S_p$ )-モノマーからは ( $R_p$ )-UpsU がそれぞれ 99:1 以上の高い立体選択性で得られることを確認した。また、all-( $S_p$ )-AGCU のホスホロチオエート 4 量体の合成を行い、平均カップリング収率が約 98% となり、TBDMS 基を有するモノマーを用いた場合（約 77%）と比較して、高いカップリング収率を示した。さらに、ホスホロチオエート 12 量体を高収率で合成することにも成功している。以上の結果より、我々が開発した CEM 基は非天然型修飾 RNA の合成においても非常に有用な保護基であることを証明した。今後は、合成したホスホロチオエート 12 量体の  $R_p$  体、 $S_p$  体それぞれの酵素耐性を確認する予定であり、その結果が出次第、特許出願および論文投稿を行う予定にしている。

スキーム 4



### (3) ボラノホスフェート RNA の合成研究

非天然型修飾 RNA の合成の一環として、核酸のリン原子上にホウ素を導入したボラノホスフェート RNA の合成研究を行った。ホウ素を導入した RNA では、次のような特徴や性質を獲得できると期待できる。

- 1) 化学的安定性の向上（リン酸結合の化学的修飾による）
- 2) 酵素耐性の向上（リン酸結合の化学的修飾による）
- 3) 細胞膜透過性の向上（ホウ素原子の疎水性による）
- 4) 低毒性（ホウ素原子は生体内に存在する原子である）
- 5) ホウ素中性子捕捉療法への適応（ホウ素の核分裂によるがん放射線治療）
- 6) 高 siRNA 活性

我々が開発した 2' 位水酸基の保護基である 2-cyanoethoxymethyl (CEM) 基を用い、ボラノホスホトリエステル法により、ボラノホスフェートアデノシンダイマー ( $\text{Ap}^{\text{b}}\text{A}$ ) を合成することに成功した。合成したダイマーはリン原子にキラリティーを有するが、立体異性体を分割することにより、光学的に純粋なボラノホスフェート RNA ダイマーを単離することに成功した。さらにこれら光学的に純粋なダイマーが以下のような特性を持つことを明らかにすることができた。

合成したボラノホスフェートダイマー  $\text{Ap}^{\text{b}}\text{A}$  は 2 分子の poly(U) と三本鎖複合体を形成した。その融解曲線を解析した結果、 $\text{S}_\text{p}$  体と推定されるボラノホスフェート  $\text{Ap}^{\text{b}}\text{A}$  は天然型である  $\text{ApA}$  より poly(U) に対して高い親和性を示し、 $\text{R}_\text{p}$  体と推定される  $\text{Ap}^{\text{b}}\text{A}$  は天然型より低い親和性を示すことが明らかとなった (図 12)。一方、poly(dT) に対するこれらダイマーの親和性は同程度であった。また、 $\text{S}_\text{p}$  体、 $\text{R}_\text{p}$  体どちらの異性体もヘビ毒ホスホジエステラーゼやヌクレアーゼ P1 に対して高い酵素耐性を示した。これら結果は、ボラノホスフェートダイマー  $\text{Ap}^{\text{b}}\text{A}$  はオリゴヌクレオチドに組み込むための有用な合成シントンの可能性を有していることを示している。特に、 $\text{S}_\text{p}$  体と推定されるボラノホスフェートを有するオリゴヌクレオチドは、天然型のオリゴヌクレオチドより RNA に対して高い親和性を示し、高い特異性が期待できることから、基礎研究および診断の分野において機能性 RNA の単離や検出へ応用が期待される。これらの成果について、論文発表、及び学会発表を行った。

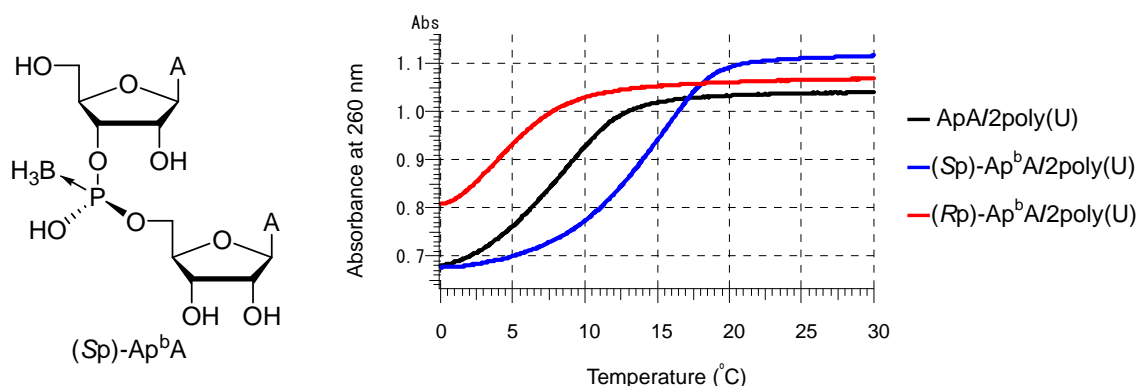


図 1 2. Ap<sup>b</sup>A と poly(U) との融解曲線

今後は、東京大学と日本新薬で共同研究契約を締結し、立体選択的な新規修飾核酸合成法の開発研究を継続したいと考えている。その一例として、A 以外の塩基でもボラノホスフェートダイマーが合成できることを確認した後、このボラノ修飾を RNA オリゴマーへ導入し、他の修飾体を上回る生物活性を示す RNA オリゴマーを開発していきたいと考えている。

### 2. 2. 3. 5. まとめ

「CEM アミダイト合成法の確立」により原料アミダイトの大量合成が可能となったことから、H19 年度より CEM 法による研究用試薬としての RNA 合成事業を開始した。さらに、核酸医薬品の原料合成を目指して、グラムスケールでの核酸製造法を確立し、核酸製造メーカーに対して新規 RNA 合成法の技術移管を行った。「長鎖機能性 RNA の化学合成」により新規 RNA 合成法は、合成純度、収率において従来法よりも優れており、鎖長が 100 程度の長鎖 RNA 合成が可能であることを実証した。「人工 mRNA の合成」は、広義の機能性 RNA 解析のためには不可欠である Cap 構造を保持する RNA の合成を実現し、また、ペプチド発現や抗ウイルス作用への応用の可能性を示した。また、その過程で新規ジリン酸化反応法を開発した。「非天然型修飾 RNA の合成研究」では、通常、立体異性体の混合物となるリン原子上に修飾を加えた RNA の光学的に純粋な修飾体合成を行った。光学的に純粋な化合物は、単一の化合物であることから、医薬品として考えた場合、活性本体の純品であり、活性および規格の面で有利である。

本プロジェクトにおいて、我々は国内 6 件、海外 2 件、合計 8 件の特許申請を行った。今後 CEM 法は、長鎖 RNA の世界的な標準合成法の一つとなることが期待できる。長鎖 RNA、修飾 RNA の需要は、世界の RNA 研究が進展するにつれて年々増加しており、試薬、診断薬、および核酸医薬品への利用が考えられる。これらの成果は、核酸を医薬品として応用する上での基盤技術として産業発展に役立つものと考えている。

**2. 2. 3. 6. 参考文献**

- 1) Ohgi, T., Masutomi, Y., Ishiyama, K., Kitagawa, H., Shiba, Y. and Yano, J. (2005) A new RNA synthetic method with a 2'-O-(2-cyanoethoxymethyl) protecting group. *Org. Lett.*, **7**, 3477-3480.
- 2) Yekta, S., Shih, I.-H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594-596.
- 3) Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140-D144.
- 4) Ohkubo, A., Aoki, K., Seio, K. and Sekine, M. (2004) A new approach for pyrophosphate bond formation starting from phosphoramidite derivatives by use of 6-trifluoromethyl-1-hydroxybenzotriazole-mediated O-N phosphoryl migration. *Tetrahedron Lett.*, **45**, 979-982
- 5) Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125-8148
- 6) Jenny, A., Hachet, O., Závorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdélyi, M. and Ephrussi, A. (2006) A translation-independent role of oskar RNA in early *Drosophila* oogenesis. *Development*, **133**, 2827-2833
- 7) Oka, N., Yamamoto, M., Sato, T., Wada, T. (2008) Solid-phase Synthesis of Stereoregular Oligodeoxyribonucleoside Phosphorothioates Using Bicyclic Oxazaphospholidine Derivatives as Monomer Units. *J. Am. Chem. Soc.*, **130**, 16031-16037