【平成２０年８月末までの研究成果】

転写制御ネットワークを解明するゲノムネットワークプロジェクトの中核機関として、プロジェクト全体の基盤リソース（転写開始点などの基盤データや、cDNAクローン等の研究ツール）の整備を実施し、コンソーシアム全体に提供するとともに、縦軸機関の研究を推進するゲノムワイドなデータ産出のサービスを実施した。また大多数のコンソーシアムメンバーが参加する発現クラスターワークショップを執り行い、コンソーシアムメンバーの連携による転写制御ネットワークの解析技術を開発した。その結果、プロモーターの活性を指標にした転写制御ネットワークをゲノムワイドに描出する解析パイプラインを構築した。これらの成果は、CAGE法などわれわれ独自の技術と、コンソーシアムメンバーの緊密な連携のもと達成されたものであり、世界的に見ても、遺伝子発現のネットワークを制御レベルで記述する、きわめて先進的な成果である。

以下に、３項目に集約し成果を記述する（１．転写制御ネットワークの解析とパイプラインの構築、２．コンソーシアムへの貢献、３．技術開発）。

１．転写制御ネットワークの解析とパイプラインの構築

さまざまなゲノムワイドの解析手法を駆使し、転写制御ネットワークの要素（ノード）と相互作用関係（エッジ）の解明を進め（静的ネットワーク）、さらに転写開始点の解析より予測されるプロモーター部分の解析から、モチーフベースの転写関連遺伝子による制御ネットワークを描いた。（動的ネットワーク）

1.1. ＲＮＡ大陸の発見（ノードとエッジの発見）とプロモーターに関する新知見（ノードの発見）

哺乳類ゲノムにおける転写開始部位および終結部位の包括的な探査から、181,047個の転写産物の5'境界ならびに3'境界を同定した。全mRNAの広汎な解析により、ゲノムの大部分（ゲノムの72%）が両方の鎖から転写物を生成させることができることがわかった。また、アンチセンス転写が哺乳類の転写調節に寄与していることを実験的に示した。

さらに、ゲノムワイドなCAGE解析により、プロモーターを、ヒトで190,513個、マウスで236,498個検出し、１つの遺伝子が平均5個のプロモーターを有することを示した。それによって遺伝子がプロモーターを使い分けることによる発現制御を行っていることが明らかになった。

1.2.プロモーターベースの世界初の動的転写制御ネットワーク解析

データードリブン、解析対象を限定する、実験材料の品質を同じにする、固有な技術を使う等の明確な研究方針のもと、ヒトモノサイト様細胞株（THP-1）の分化を題材として、遺伝子発現、転写開始点、クロマチン情報など総合的なデータを産出した。これはデータ間の相互比較を可能にする、繊細な実験デザインにより産出された、世界初のデータセットである。とりわけ、転写開始点を１ベース精度で同定するため、また、プロモーター活性を定量的に測定するため、DeepCAGE法を確立し、経時変化を追跡することで転写因子と被制御遺伝子のカスケード（転写制御の動的ネットワーク）の構築を目指した。

発現クラスターワークショップを立ち上げ、遺伝子発現制御ネットワーク解析のための一連の手法（データ産生、データベース化、アノテーション／解析）の確立をめざした。クラスターワークショップは、多数のコンソーシアムメンバー（37研究室、148名）が参加しており、連携を重視するＧＮＰを象徴する活動となっている。

前述の産出されたデータセットを解析して、ゲノムワイドに動的転写制御ネットワーク解析の手法を確立することができた（図１）。すなわち、DeepCAGE情報から、プロモーター位置と活性を同定し、プロモーター上に存在する転写因子結合モチーフの活性（モチーフアクティビティ）を情報科学的に推測することにより、プロモーター発現のモデル化を行った。さらに、このモデルをベースとして、プロモーターベースで世界初の動的転写制御ネットワークを描き、その妥当性を既報、ChIP-chip実験, siRNAによるパータベーション情報を用いて検証し、確認した（図２）。これによって、分化に代表される細胞変化は、少数のマスター遺伝子によって担われているのではなく、多くの転写因子が協調的に役割を果たすことにより達成されていることが明らかとなった（図３）。論文は現在投稿中である。

取得したデータや解析データは、プラットフォームを通じてコンソーシアムに随時開示した。また、遺伝子間の制御関係情報を集積したエッジデータベースを作成した（図４）。

さらに細胞の分化や転写制御に関与する機能性 RNA に関する興味深い知見が得られている。短鎖の RNA（shortRNA）の解析からは、転写開始点付近にマッピングできる約 18 ベースの shortRNA を同定し、それらが、GC 配列が多く含まれているプロモーター部分に相補的な配列を有することから、転写制御への関与が示唆されている。

また、miRNA 生成に関与する酵素群の発現解析から、分化の初期において AGO２, AGO３の発現レベルが一過性に昂進することが観察されていることと、特定の miRNA の強制発現が THP-1 の分化を促進する効果のあることが観察されている。

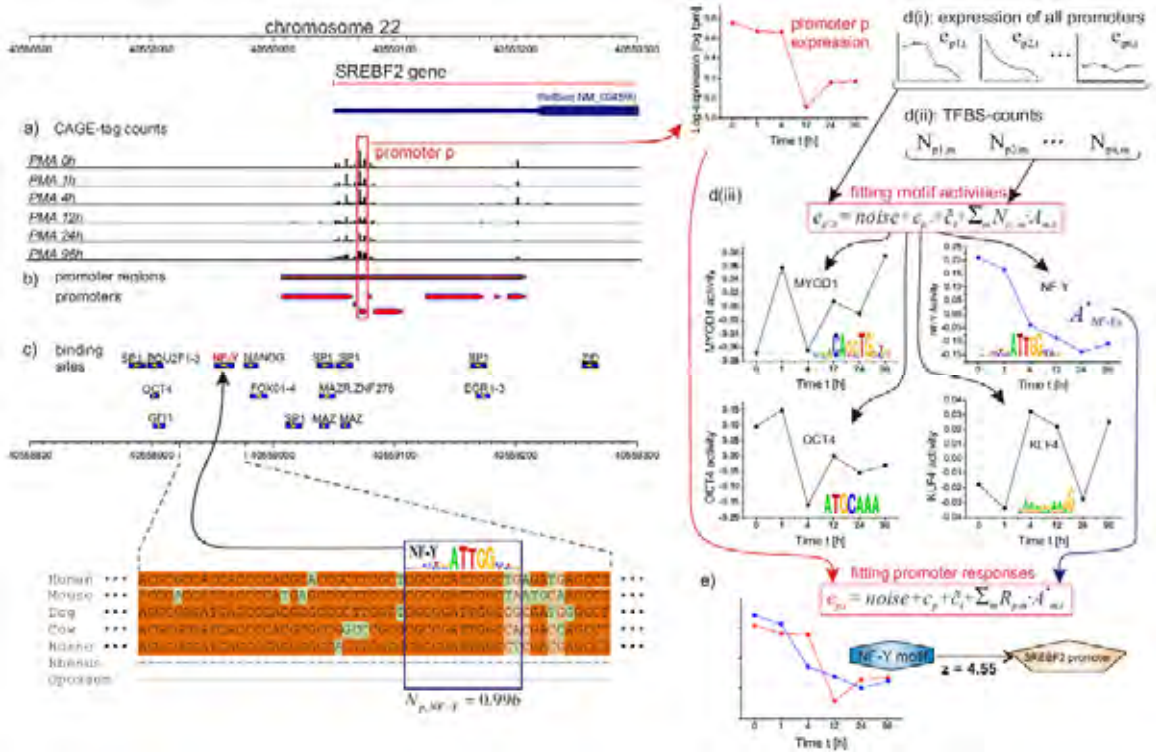

図１：動的転写制御ネットワークの解析戦略。DeepCAGE によりプロモーターを同定し(a, b)、プロモーター内の転写因子結合部位（モチーフ）を同定する(c)。すべてのプロモーターの発現情報から各モチーフアクティビティを計算し(d)、各プロモーターを制御するモチーフを推測する(e)。

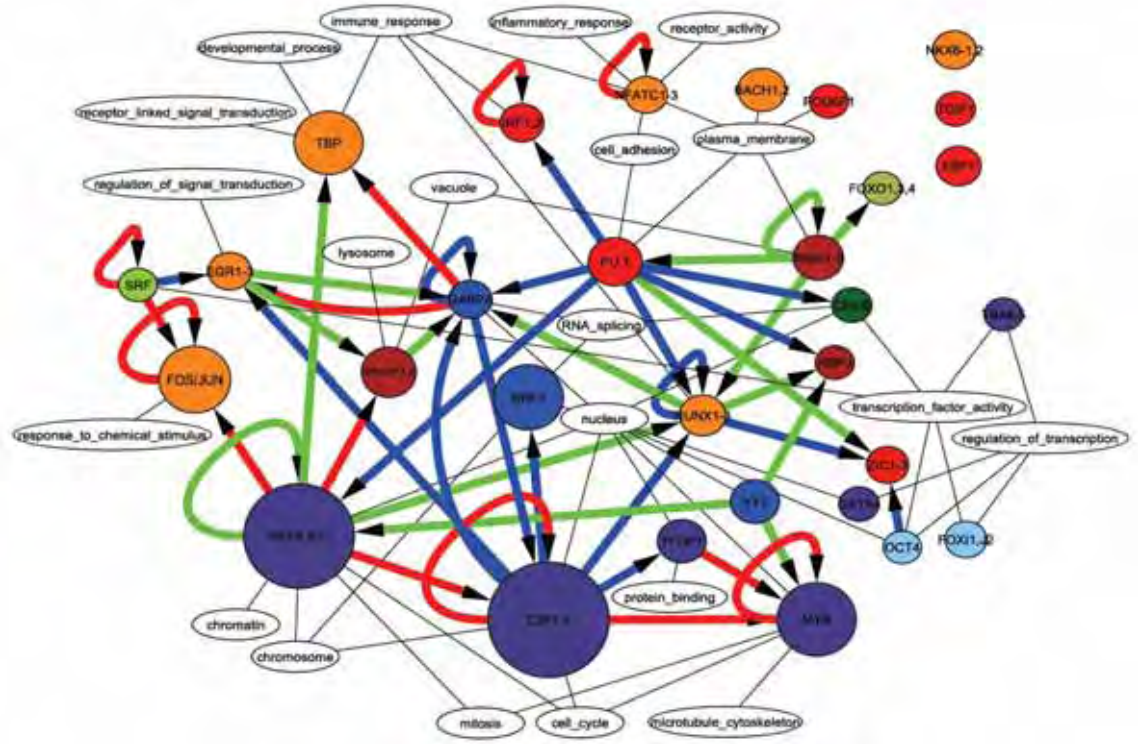

図２：モチーフベースのコア転写制御ネットワーク。各モチーフアクティビティをクラスター解析し、分化に応じて活性が更新するものを赤系統で、減少するものを青系統で示した。矢印は制御を示しており、エッジの色は文献（赤）、ChIP-chip（青）、siRNA 実験（緑）によりサポートされていることを示している。さらに、モチーフアクティビティが制御すると推定される遺伝子群で有意に観察されるオントロジーを楕円内に示した。

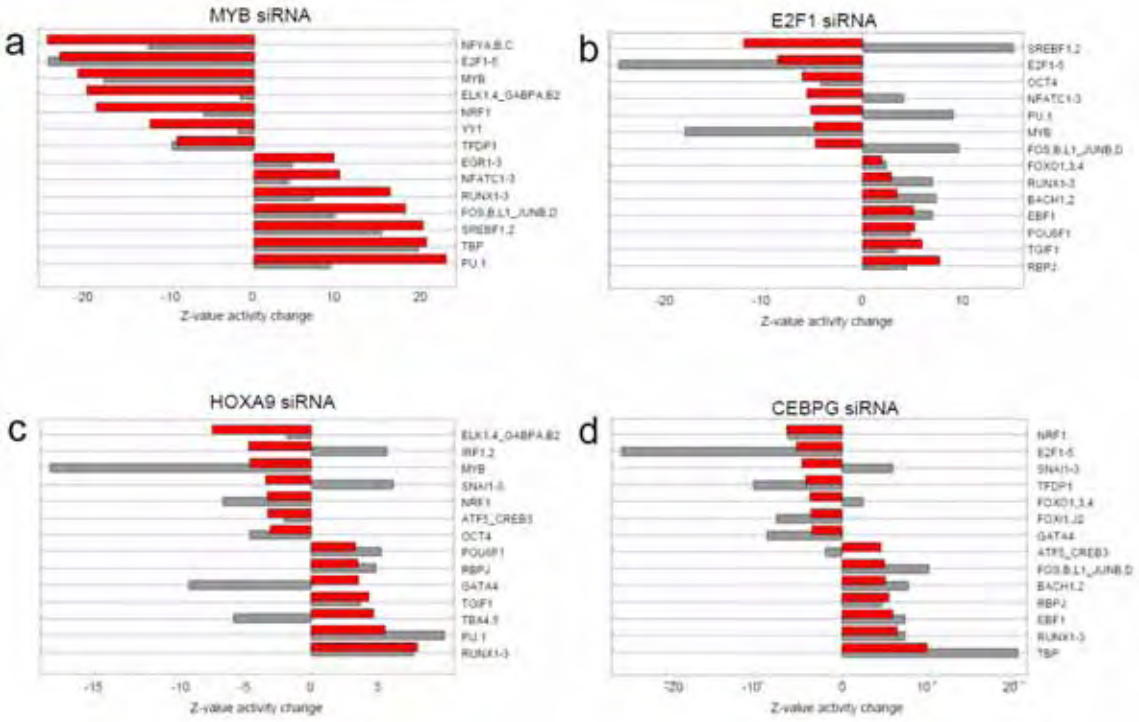

図３：転写因子ノックダウンによる各モチーフアクティビティの変化。24 実験のうち 4 つを示した。各転写因子ノックダウンによるモチーフアクティビティの変化の度合い（赤色バー）を、PMA 刺激による分化に伴う変化の度合い（灰色バー）と比較した。各転写因子は一般的に PMA とよく似た変化をする。その度合いは MYB では高いが、他の転写因子では PMA と異なる活性変化も観察された。この結果から、多くの転写因子の協調的作用が分化に必須であることが明らかとなった。

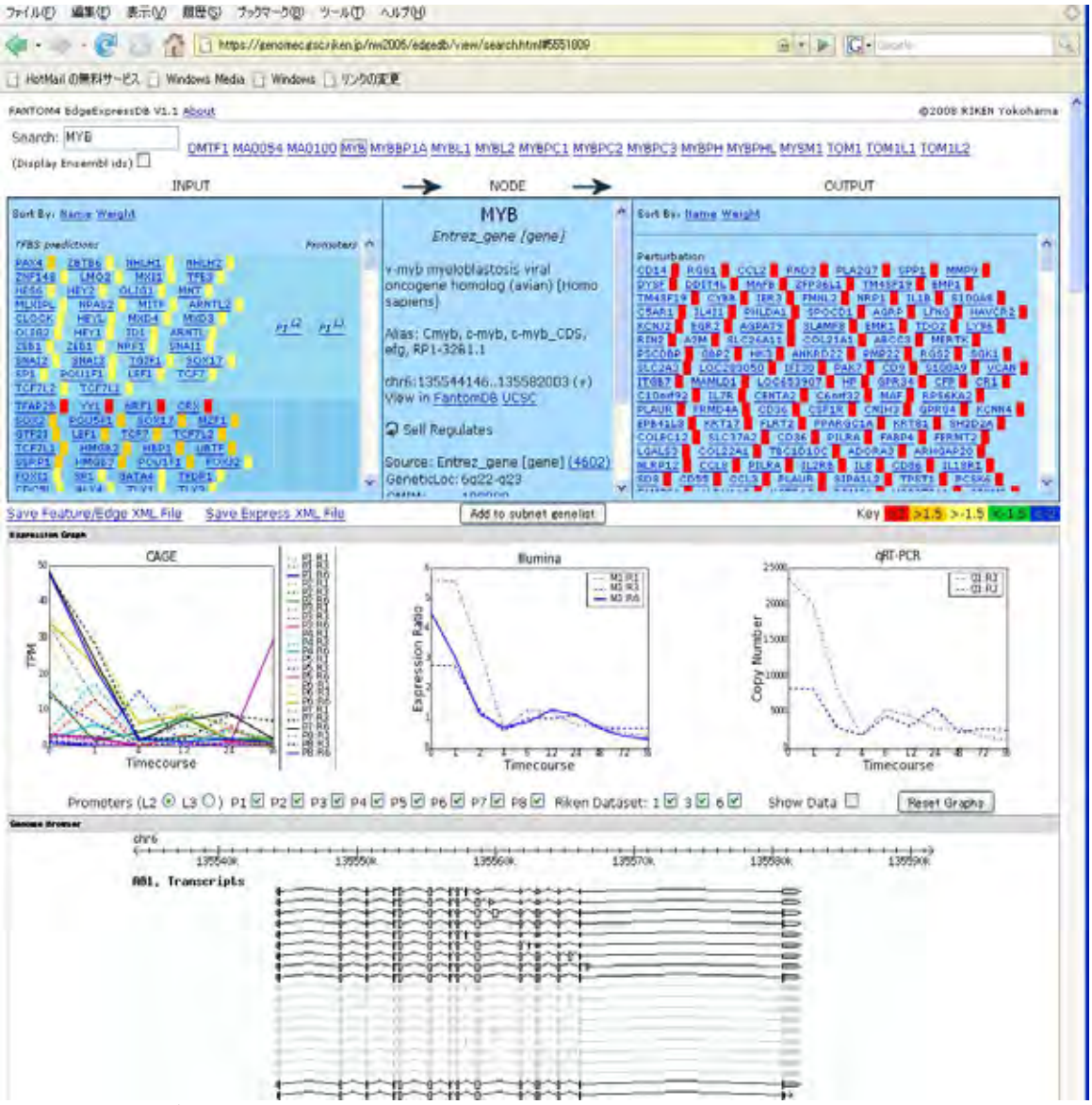

図４：エッジデータベース。遺伝子制御についての推測情報を確度とサポート情報と共に閲覧できるシステムになっている。また、遺伝子発現情報が示されており、ゲノムビューアーともリンクしている。

２．コンソーシアムへの貢献

中核機関として、ゲノムネットワーク研究の推進に資する基盤データの産出、cDNA クローン等のコンソーシアムメンバーへの配布、および縦軸研究機関からの依頼によるデータ産出を実施した。また実施会議議長のもと、発現クラスターワークショップを実施し、コンソーシアムメンバーの連携推進に貢献した。

2.1 クローンリソースの整備とコンソーシアムメンバーへの提供

cDNA 等のクローンリソースを、コンソーシアムメンバーのリクエストに従って配布した。配布件数は約 120 件、ヒト cDNA クローンは約 28 万クローン、マウス c DNA クローンは約 900 クローン、siRNA クローンは２セット、3,000 クローン、および 23 細胞株を配布した。

・完全長 cDNA クローン（6,825 遺伝子）
ヒトの転写関連遺伝子を中心として、核内タンパク、シグナル伝達分子、レセプター・リガンド分子に関し完全長 cDNA クローンを整備した。
・ORF エントリークローン（3,933 遺伝子）
転写関連遺伝子などの ORF エントリークローンを整備した。
・転写関連遺伝子 500 個に対する siRNA 発現クローン（1,500 クローン）
活性評価を付与したsiRNA発現クローン（転写制御因子500個に対する1,500クローン）を作製した。

2.2 横軸データの産出およびコンソーシアムへの提供
下記の計測データを、ゲノムネットワークプラットフォームよりコンソーシアムに開示した。中でもヒトモノサイト様細胞株ＴＨＰ-1 の分化過程で採取したデータは、世代管理された試料を使用し、統一した考えで系統的に実験を行い、相互に比較することが可能な、画期的なデータセットである。
・ｑRT-PCR による転写制御因子（2,315 遺伝子）の発現情報
ｑRT-PCR により、ヒトの主要な 41 組織と 27 細胞株と、THP-1 の分化過程の 10 タイムポイント
・CAGE 法による転写開始点情報
ヒト 69 組織、48 細胞株および THP-1 の分化過程の 6 タイムポイントデータ（43,125,302 タグ）。
・短鎖 RNA の発現情報
THP-1 の分化･活性化過程の 11 タイムポイント､およびヒト肝がん由来細胞株 HepG2 の短鎖 RNA の発現情報（合計 13,973,634 シーケンス）。
・転写関連因子タンパクの相互作用（PPI）情報
ヒトおよびマウス転写関連因子の M2H 法によるタンパクタンパク相互作用情報と外部データベースに登録されている相互作用の情報を解析し、7,011 組の相互作用のある遺伝子をリストアップした。
・タンパク－ＤＮＡ相互作用情報およびクロマチン構造情報
ChIP-chip 法により、THP-1 細胞のヒストン H3 アセチル化情報の解析を行った。さらに、転写因子 SPI1 と SP1 の結合情報、及び転写制御にかかわる RNA ポリメラーゼⅡの情報を得た。
・マイクロアレイによる遺伝子発現情報
THP-1 について、分化過程の 10 タイムポイント、および 54 遺伝子のノックダウンや 12 種類の miRNA 前駆体の強制発現によるパータベーション後の遺伝子発現情報、また、miRNA マイクロアレイよる THP-1 の miRNA（534 遺伝子）の発現情報を得た。

2.3.依頼解析
縦軸研究機関の要請に応じ、CAGE 等のデータ産出を実施し、縦軸研究の推進に貢献した。
・CAGE 7件、shortRNA 3件、PPI 3件、qRT-PCR ３件。詳細は、【ＧＮＰとの関連・ＧＮＰへの貢献】を参照のこと。

３．技術開発
プロモーターベースの転写制御ネットワークの解析に必須の技術の開発、および、より定性的あるいは定量的に転写制御ネットワークを描出する技術を開発した。

・CAGEバーコード技術の開発
転写制御ネットワークの解析には、複数のサンプル間におけるプロモーター活性の定量的な比較が必須である。転写開始点を定量的に同定するCAGE法は、そのために利用できる唯一の方法であるが、従来は複数のサンプルを独立に調製しなくてはならず、実験間のばらつきのためサンプル間の比較が容易ではなかった。そこでサンプルごとに数ベースのオリゴヌクレオチドをCAGEタグに付加するバーコード技術を開発し、サンプル間の定量的な比較を可能とした。

・転写制御ネットワークをより定量的に解析するため、また、データ精度の向上に資するため、次世代型シーケンサー（Solexaシーケンサーや SOLiDシーケンサー）を立ち上げ、deepCAGE法の最適化を行った。現在Solexaでは1ランで約1,400万シーケンス以上読むことができ、SOLiDでは約2,000万シーケンスを得ることができるようになった。現在CAGEタグシーケンスのゲノムへのマッピング率はそれぞれ約50％である。SOLiDのゲノムシーケンスのマッピング率は80％以上である。

・CAGE法はプロモーター活性を計測できる唯一の重要な技術であるが、多数の工数を必要とするCAGEライブラリ作製プロトコールは、CAGE法を広く普及させる上でひとつの弱点であった。そこでCAGEライブラリ作製事業を実施している機関にアイデアを提供し、工程の簡素化、スピードアップ化を図った。複数回必要なDNA精製工程を簡便化し、ライブラリ作成の工程時間を約３分の１に短縮した。

【ＧＮＰとの関連・ＧＮＰへの貢献】

ａ）転写・制御の解明に関連する研究成果
転写開始情報をもとにモチーフアクティビティの値を計算し、モチーフベースの転写制御ネットワークを描いた。

ｂ）縦軸研究（または横軸研究）への貢献に繋がる研究成果
１．下記の依頼解析を実施した。
［CAGE解析］
　　東京大学 井上聡先生 17サンプル
　　東京医科歯科大 高柳先生 2サンプル
　　埼玉医科大 岡崎先生 7サンプル
　　国立成育医療センター研究所 浅原先生 4サンプル
　　大阪バイオサイエンス研究所 裏出先生 6サンプル
［shortRNA解析］
　　東京大学 井上聡先生 17サンプル
［PPI］
　　10機関より 229件
　　福岡大学 白澤先生 4クローン
　　東京大学 井上聡先生 15クローン
　　九州大学 福井先生 1クローン（4分割）
［qRT-PCR解析］
　　東京大学 井上聡先生 17サンプル
２．クローン配布
　平成20年度までにコンソーシアムから約120件の依頼があり、約28万個のヒトｃDNAクローン、約900個のマウスcDNAクローン、2セット3000個のsiRNA発現クローン、23個の細胞株を配布した。
３．データベース講習会を遺伝研と共同開催した。（2006年2月、2006年7月）

【当該研究に関して活用した研究データ・リソース】

THP-1から得られた遺伝子転写開始点情報（deepCAGE）、短鎖のRNAの発現情報、マイクロアレイを用いた遺伝子発現情報、タンパク-DNA相互作用情報、クロマチン情報、タンパクタンパク相互作用情報、qRT-PCRによる遺伝子発現情報

【主な研究論文とその概要】

1 P. Carninci T. Kasukawa , S. Katayama ,他 , The transcriptional l,scape of the mammalian genome, *Science*, **309**, 1559-1563 (2005)
RNA 大陸の発見。トランスクリプトの 50%以上が ncRNA を見出した。

2 S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engstrom, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, and C. Wahlestedt, Antisense transcription in the mammalian transcriptome, *Science*, **309**, 1564-1566 (2005)
トランスクリプトが生成するセンス／アンチセンスペアによる転写制御機構の発見。

3 P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A. Semple, M. S. Taylor, P. G. Engstrom, M. C. Frith, A. R. Forrest, W. B. Alkema, S. L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S. M. Grimmond, C. A. Wells, V. Orlando, C. Wahlestedt, E. T. Liu, M. Harbers, J. Kawai, V. B. Bajic, D. A. Hume, and Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution, *Nat Genet*, **38**, 626-635 (2006)
哺乳動物のプロモーターを、新規技術；CAGE 法により網羅的に同定した。

4 R. Kodzius, M. Kojima, H. Nishiyori, M. Nakamura, S. Fukuda, M. Tagami, D. Sasaki, K. Imamura, C. Kai, M. Harbers, Y. Hayashizaki, and P. Carninci, CAGE: cap analysis of gene expression, *Nat Methods*, **3**, 211-222 (2006)
CAGE 法を開発した。

5 N., Maeda, Nishiyori H, Nakamura M, Kawazu C, Murata M, Sano H, Hayashida K, Fukuda S, Tagami M, Hasegawa A, Murakami K, Schroder K, Irvine K, Hume D, Hayashizaki Y, Carninci P, Suzuki H.Development of a DNA barcode tagging method for monitoring dynamic changes in gene expression by using an ultra high-throughput sequencer. Biotechniques. 45, 95-7 (2008)
CAGE バーコード技術の開発

他、合計で、英文論文;82報、和文総説；5報、国内外学会等発表；125件

【特許出願数の実績】

出願済

・特願2005-362337
発明の名称：ⅡＳ型制限酵素を用いる翻訳終始コドンの除去方法
発明者：林﨑良英、鈴木治和、伊藤昌可

・特願2008-050744
発明の名称：酵素の反応性を改善する方法
発明者：林﨑良英、小島美樹、Carninci Piero、Harbers Matthias

出願準備中
発明の名称：転写因子群の活性を予測する手法
発明者：林﨑良英、鈴木治和、Erik van Nimwegen

出願準備中
発明の名称：ノックダウンCAGE法
発明者：林﨑良英、鈴木治和、Alistair Forrest

ゲノムネットワークプロジェクトの
課題の報告

プログラム名	（1）ゲノム機能情報の解析　（2）次世代ゲノム解析技術の開発 （3）個別生命機能の解析　（4）ヒトゲノムネットワークプラットフォームの構築（5）動的ネットワーク解析技術開発					
課題の分類	（1）中核機関　（2）指定課題　（3）公募課題					
課　題　名 （和英併記）	ゲノムネットワーク解析に向けたヒトcDNAクローンの整備 Collection of cDNA Clones for Genome Network Analysis					
代表研究者名 （所属機関・職名）	菅野　純夫（国立大学法人東京大学・教授）					
年度別研究費 （千円）	１６年度	１７年度	１８年度	１９年度	２０年度	合　計
	263,000	263,000	0	0	0	526,000

		研究者名	所属機関・部局・職	専門分野
研究組織図	代表機関	☆菅野純夫	東京大学・大学院新領域創成科学 研究科・教授	分子生物学 ゲノム医科学
	分担機関			
		※ 研究者名欄は、適宜増減してください。		

【研究目的】

タンパク質をコードすると考えられるヒト遺伝子（ensambl 遺伝子を基準とする）のほぼ全てを、理化学研究所ゲノム科学総合研究センターと緊密に連携しつつ、２年間で、転写因子でない遺伝子を中心にヒト cDNA を収集する。また、ゲノムネットワーク解析の効率化をはかるため、多目的ベクター系へのｃDNA の再クローン化を行う。cDNA クローン収集データベースを作成運用していく。

【課題の概要】

（1）cDNAクローン収集データベース構築・維持：ヒトcDNAの配列情報と、そのクローン化進行状況を同時に把握できるデータベースを構築・維持する。

（2）事故クローンの修復：大きな欠失などがあるものから修復する方針とした。

（3）全長配列解析とクローンの収集：完全長のcDNAが得られていないものでcDNAを分離する必要があるものについては、5'端の部分配列情報のあるクローンから候補クローンを選定し、その全長配列決定を行うのが効率がよい。5'端配列情報のあるクローンは、菅野グループが保持するもの、林崎グループが保持するもの、日本企業より入手可能なものを合わせ、5'端配列情報のあるクローン１５０万に、企業より入手可能なもの若干数を加え、その中から上記のｃDNA のデータベースを利用し全長配列決定を行ってクローンを選ぶ。

（4）新規クローンの分離：遺伝子としての登録があるにもかかわらず、世界中のデータベースにESTが存在しないものの分離。

（5）cDNAの多目的ベクター系への再クローン化：ORFをGATEWAY多目的ベクター系に再クローン化して、ゲノムネットワーク研究のために供給する。

【課題終了時までの研究成果】
本研究では、基準となる ensambl 遺伝子２２２３７種中、割り当てを理研７１７１、東大１５０６６とし、東大割り当て１５０３６種中、NEDO・野村プロジェクトでカバーされている５５７９遺伝子を除き、残り９４８７種について、クローン収集と GATEWAY 化を目指してきた。ただ、EST が存在しないクローンの入手困難さや予算の制約等により GATEWAY 化目標は、当初の８０００種より、減らし EST が存在しない新規クローンの分離に注力することとした。
この結果、１６年度と上記の１７年度の活動を合わせ、ｃDNA クローン収集済みのものが８４７２種、ｃDNA クローン収集できなかったものが１０１５種であり、達成率は８９％であった。また、収集したもののうち６０８７クローンを GATEWAY 化した。
作成したデータベースについては理化学研究所と共同で利用し、計画の進行状況を随時把握するとともに、収集データを国立遺伝学研究所の五條堀らに供給し、ヒトゲノムネットワークプラットフォームへの連携もはかった。

【ＧＮＰとの関連・ＧＮＰへの貢献】
ａ）転写・制御の解明に関連する研究成果
本研究は、横軸研究の中でも、特に基盤的なリソースであるcDNAの収集整備を目指したものであった。

ｂ）縦軸研究（又は横軸研究）への貢献に繋がる研究成果
収集したｃDNAクローンは、理研を中心に、GNP参加グループに配布を行った。東大からも、FLJクローンについては、慶應義塾大学理工学部生命情報学科や国立成育医療センター研究所のグループへ配布を行った。なお、整備されたｃDNA リソースやデータベースはゲノムネットワークプロジェクトに止まらず、広範な医学・生物学研究において不可欠の研究リソースであり、医学・生物学研究全般の加速に多大な貢献を行うものと考えられる。

【当該研究に関して活用した研究データ・リソース】
FLJｃDNAクローン、MGCcDNAクローン。（意味が違ったら削除お願いいたします）

【主な研究論文とその概要】
なし

【特許出願数の実績】
なし