

平成 29 年 5 月 31 日現在

機関番号：82401

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26640120

研究課題名（和文）1細胞マルチオミクス技術の開発

研究課題名（英文）Technical development of single-cell multi-omics

研究代表者

笹川 洋平（Sasagawa, Yohei）

国立研究開発法人理化学研究所・情報基盤センター・上級センター研究員

研究者番号：10404344

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本課題では、1細胞からエピゲノム情報とRNA情報を取り出すことが出来るマルチオミクス技術の開発を目指した。課題推進中に、提案した方法の一部が達成されてしまったが、途中で方向転換して、新生RNAを含むポリアデニル化されていないRNA分子も捉えられるsingle-cell RNA-seq法を確立することに成功した。新生RNAを検出することで、ChIP-seqなどで検出される転写調節要素の、活性化状態を知ることが出来ると期待される。また、cell barcoding技術をランダムプライマーに持ち込むことを検討し、高出力化への道筋をつけた。

研究成果の概要（英文）：In this study, we tried to develop technology for simultaneous detection of epigenome and transcriptome from one cell. We separated cytosol and nucleus from a single-cell using pipet tip. We tried to use nucleus for detection of epigenome. Moreover, we succeeded to develop single-cell RNA-seq method, which detect non-poly-A RNA including nascent RNA. Moreover, we tried to develop cell-barcoding method using random primer for high-throughput detection of RNA.

研究分野：バイオテクノロジー

キーワード：1細胞RNA-seq 1細胞エピゲノムシーケンス

## 1. 研究開始当初の背景

ゲノム DNA のクロマチン構造やメチル化などのエピゲノムの変化が及ぼす RNA 量への影響は未だ理解されていない部分が多い。一般的に、エピゲノム情報と RNA 情報は、それぞれ細胞集団を用いて計測するため、1 細胞内でのエピゲノム情報と RNA 情報のリンクが切れてしまい、平均化された情報の相関解析になる。例えば、我々の体の臓器や組織を構成する細胞集団はゲノム情報が同一にも関わらず、分化状態や細胞状態に応じてエピゲノム状態や RNA 量の不均一性が観察される。つまり細胞集団での解析では、分化状態や細胞状態の混じった細胞集団の平均の解析となる。1 細胞から両者を同時に計測することができれば、分化状態や細胞状態をわけてから、両者の相関解析を精緻化することも出来ると考えられる。

我々が開発した Quartz-Seq をはじめ、1 細胞 RNA-seq 法は、RNA 量の不均一性を計測することで、分化状態の異なる細胞をクラスター分けし、そのマーカー遺伝子を取得することが可能になってきた。一方で、1 細胞 RNA-seq だけでは、各クラスターの RNA レベルのマーカー遺伝子を見つけることが出来ても、その変化を説明できるエピゲノムの変数候補を得ることは出来ず、両者の関連の理解が進まない。その為、1 細胞から両者を計測する 1 細胞マルチオミックス技術の開発が望まれていた。

## 2. 研究の目的

本課題では、1 細胞から RNA 情報とエピゲノム情報を取り出せる、またその代替になる多次元情報を 1 細胞から取り出せる 1 細胞マルチオミックス技術の確立およびそれに資する重要な分子生物学的な反応の確立を目的とした。課題では、主に実験技術の開発に注力した。

## 3. 研究の方法

### (1) エピゲノム情報と RNA 情報を物理的に分ける方法

Mouse embryonic stem cells を様々な濃度の界面活性剤 (0.05%-1%) を含む緩衝液で処理し、1 つの核の total RNA 量を計測し、核へのダメージを検証した。また多数の細胞由来の細胞質および核から抽出された total RNA を用いて RNA-seq を行い、発現定量解析を行った。

### (2) Non-poly-A RNA の検出

まず mouse embryonic stem cell 由来の精製した total RNA から希釈することで 10pg total RNA を作製し、single-cell RNA-seq の性能検証に使用した。ランダムプライマーから rRNA の配列と完全に一致する配列を除いた NSR (Not so random) primers を合成し、逆転写反応に用いた。逆転写反応後、Klenow fragment など中温性 DNA polymerase を用い

て、2 本鎖合成を行った。得られた産物を、磁気ビーズで精製を行い、得られた DNA 産物を、Nextera XT キットを用いてライブラリ作製を行った。得られた sequence library DNA を、Illumina sequencer で解析し、発現定量解析を続けて行った。

### (3) Adaptor 配列付きランダムプライマーが及ぼす逆転写反応の影響

mouse embryonic stem cell 由来の精製 RNA を用いて、random primer および様々なアダプター配列のついた random primer で逆転写反応を行った。その後、*Nanog* などの遺伝子発現を qPCR によって計測し、値を比較した。

### (4) Cell barcode sequence

研究協力者の團野宏樹研究員と連携し、シーケンスレーベンシュタイン距離 (編集距離 5) およびハミング距離 (編集距離 5) によりシーケンスエラーやライブラリ作製でのエラーを補正できる配列を計算により得た。様々な cell barcode の長さで計算し、それぞれの塩基組成を計算した。

### (5) Nucleosome pattern を解析する際に得られる 1 細胞あたりの DNA 量

mouse embryonic stem cell から核を抽出し、double strand specific nuclease にて DNA 分解処理をした。その後、複数の核から DNA を抽出して、高感度 DNA 長解析装置 (Bioanalyzer High-sensitivity kit) にて解析した。

### (6) Sequence library DNA への変換効率の計測とその改善

市販の mouse genomic DNA を超音波により断片化した。断片化した genomic DNA (1 ng or 100 pg) を用いてライブラリ作製を行った。ライブラリ作製方法は、両端の配列を修復したあと、AT ライゲーションにより両端にアダプターを付与する方式である (図 1)。ライブラリ作製に用いたインプット DNA 量と作製後のライブラリ DNA 量を qPCR などで計測し、両端にアダプターが付与された効率を計測した。ライブラリ作製においては、KAPA Hyper Prep Kit を用いた。我々が最終的に最適化で得られた方法では、以下の改変を行った。KAPA Hyper Prep Kit の反応ボリュームを 1/5-1/10 に、End repair では 37°C 1 時間反応した。

### (7) Single-cell ATAC-seq の実施

上記条件にて核を抽出して、open chromatin を解析することの出来る Single-cell ATAC-seq を行った。オリジナル論文では、微細流路中にて反応を行っており、同時測定が目的である我々の用途に合わない。mouse embryonic stem cell から抽出した核を、SH800 にて 96 well plate に 1 つずつ分取し、オリジナルの論文 (Nature 523, 486-490) を

参考にライブラリ作製を行った。得られたシーケンスライブラリ DNA は Bioanalyzer にて解析した後、Illumina sequencer で解析した。

### (8) 実験環境中に浮遊する DNA コンタミネーションの影響とその対策

実験環境中に浮遊する DNA コンタミネーションの影響を計測するために、まずライブラリ作製に使用する KAPA Hyper Prep Kit を 2 つ用意して、それぞれ清浄度 ISO class 1 のクリーンベンチと、清浄度 ISO class 6-7 の通常実験室で、2-3 回ライブラリ作製を行い、コンタミネーションのリスクに晒した。クリーン度はパーティクルカウンターで計測して確認した。その後、2 pg の断片化ゲノム DNA と non-template control (0 pg) を用いて、清浄度 ISO class 1 でライブラリ作製した。アダプターの付与されたライブラリ DNA は、19-20 cycle 程 PCR による増幅を行った後、Bioanalyzer により DNA 長を計測した。Non-template control から得られたライブラリ DNA に関しては、illumina sequencer でシーケンス反応を行い、Basespace cloud により metagenome 解析を行った。

### (9) データ解析

研究連携者の團野宏樹研究員と連携し、Mac pro に、ハイスループット single-cell RNA-seq 法である Drop-seq の解析パイプラインを導入した (<http://mccarrolllab.com/download/922/>)。解析パイプラインは Docker 化したものを Mac pro に組み込んだ。

## 4. 研究成果

### [物理的な分画方法]

最初に、1 細胞から RNA 情報とエピゲノム情報を取り出すために、物理的な分画方法について実験した。まず 1 細胞を核と細胞質にわけられるか検証した。核をエピゲノム解析に、細胞質を RNA 解析に使うことを想定した。細胞の直径が約 10 $\mu$ m であったが、核の直径は 7 $\mu$ m 程度であった。1 細胞から得られる total RNA 量はおよそ 10 pg であったが、1 個の核あたりの total RNA 量はおよそ 3.5 pg であった。直径から考えられる体積の差は、細胞と核で 35% 程度であり、細胞全体と核の total RNA 量の比率と一致していた。界面活性剤の濃度を 0.5% 程度以上にすると、1 個の核あたりの total RNA が減少し、1 pg 程度まで減少することから、強い界面活性剤では、核からの RNA の漏れ出し方が激しくなると考えられた。核・細胞質・細胞から抽出した total RNA をつかって bulk RNA-seq 解析した結果、細胞質 (約 6pg total RNA/cell) と細胞全体の遺伝子発現プロファイルと比較した結果、限られた遺伝子を除いて、細胞質だけでも殆どのトランスクリプトを検出することが可能であり、細胞全体と細胞質の遺伝子発現プロファイルは十分似ていた。1 細胞の

細胞質を RNA 解析に使っても良いと判断した。つぎに 1 細胞を核抽出 buffer に採取した後、遠心をして、細胞質を得ることが出来るか人工的な系を使って検証した。磁気ビーズを核にみだして、上清だけ採取できるのか、ロボット (Agilent Bravo NGS workstation) をもちいて実験した。人間の手によるピペティングと異なり、ロボットを使うことで安定して上清だけ吸えることがわかり、物理的に両者を分けることは技術的に可能であることを確認した。

### [エピゲノム解析技術]

Embryonic stem cell から、個別に取得した、1 個の核を、既報の single-cell ATAC-seq を改変することで解析した (Nature 523, 486-490, 2015)。得られたシーケンスライブラリ DNA は、わずかに nucleosome pattern が見えた (図 2)。続けて sequence 解析を行った。結果としては、genomic DNA に mapping することができる reads が全体の 10-20% 程度以下であった。Mapping reads しない read はシーケンスライブラリ DNA を作製したときに使用するアダプターやプライマーなどの由来ではないことを確かめた。おそらく、予期しない外部からの微量のコンタミネーションであることが疑われた。非常に微量な DNA をターゲットとしているため、僅かな外部からのコンタミネーションにも脆弱であることが考えられた。ATAC-seq 自体は、Tn5 トランスポゼースを使用しており、全ての DNA に対して反応する。DNase をベースとする open chromatin sequence では、意図しないコンタミも同時に分解することが期待された。

次に、open chromatin を解析するため dsDNase を用いた酵素反応条件を確かめた。embryonic stem cell 由来の核を dsDNase 処理した。結果、1 細胞あたり約 2 pg の断片化されたゲノム DNA が得られ、Nucleosome pattern を示した (図 3)。

次に、これら断片化 DNA をシーケンスライブラリ DNA に変換できる効率を計測した。この効率は、1 細胞からエピゲノム情報を取り出す時の精度に直結すると考えられる。まず人工的な系を用いて、Y 型の sequence adaptor が両端に ligation される効率を計測した (図 1)。この方式は、最も一般的な Illumina 社 sequencer 用のライブラリ作製方法である。結果としては、kit の初期条件では、15% 程度の効率しか無かった。そこで、様々な条件を最適化した結果、以下 3 点が重要であることがわかった。1) 反応 volume。1/5-1/10 の volume にする 2) 末端構造を修復する温度条件を 37 $^{\circ}$ C 60 分にする。3) Y 型 sequence adaptor の濃度を、初期設定値より濃くする。これらの改善点により、最終的に 60-80% の変換効率になった。

上述の通り、様々な工程において外部からのコンタミネーションへの対策は重

要であることが示唆されていた。そのため、ligationによるライブラリDNA作製でのコンタミネーションへの影響を調べた。0pg、2pgの断片化DNAを、クリーンベンチ内・外で予め使用した試薬を使いライブラリ作製した。結果、クリーンベンチ外で試薬をつかった場合、0 pg(non-template control)において、意図しないDNAのピークが見られた(図4)。クリーンベンチ内での使用の場合、それらは見られなかった。クリーンベンチ内でライブラリを作製した場合、2 pgからコンタミネーション無く、ライブラリ作製が可能であることがわかった(図4)。これらの結果は、pgオーダーの微量DNAを想定する1細胞エピゲノム解析は、コンタミネーションの対策を必要とすることを示唆している。

ではこのコンタミネーションの由来が何だったのかを知るために、クリーンベンチ外の non-template control で得られたシーケンスライブラリDNAをIllumina sequencerで解析した。その後、metagenome解析した。結果、ヨーグルトの生産に関わる微生物や海洋系細菌などが見つかった(図5)。部屋の環境なのか、人間由来か最終的には判別がつかなかった。一方、クリーンベンチ外で、1回に起きるコンタミネーションは、0.1 pg-0.25 pg程度であることも分かった。

#### [RNAの検出]

エピゲノム情報とRNA情報の1細胞同時計測は、物理的分画を主軸とする方法により、課題推進中に幾つかの方法が提唱された。一方で、細胞集団のことを理解するためには、高出力化が未だ求められている。また、GRO-seqなど新生RNAを計測する方法では、ChIP-seqなどのエピゲノムシーケンス技術によって発見される転写調節要素の、活性化状態を知ることが出来ることが知られている。新生RNAの計測は、エピゲノム変化によるRNAの転写制御への影響を直接計測していると考えられ、エピゲノム情報とRNA情報の同時計測の代替になる可能性がある。一方で、1細胞での網羅的な計測方法は未だ報告されていない。より報告価値が高く、高出力化も同時に見込める後者について、開発を進めた。

新生RNAなどのpoly-Aが付与されていないRNAも検出するためには、逆転写反応時に使用するプライマーを、ランダムプライマーをベースする必要がある。これまでの方法では、逆転写反応時にpoly-dT配列をベースとしたプライマーを使用していたので、poly-A RNAしか検出することが出来なかった。NSR primerを用いた逆転写反応と単純な2本鎖DNA合成を行うことで、高い定量性能を出せることがわかった(図6)。一方で、新生RNAなども含む形で検出できるようになって、1細胞ごとに細胞を処理するため、高出力化には至っていない。近年の1細胞RNA-seqは、mRNAを対象とした方法として成熟しつつあり、少なくとも百以上の1細胞を1実験の

処理単位として、数千・数万の1細胞の解析が可能になってきている。高出力化により、細胞集団の理解につながっている。そこで、ランダムプライマーベースの1細胞RNA-seqを高出力化するための開発を行った。

高出力化のkey technologyとしては、cell barcoding技術がある。1細胞ごとに個別のuniqueな配列を逆転写反応で付与することで、得られたfirst-strand cDNAをまとめて精製し、その後のシーケンスライブラリ作製を高出力化することが出来る。ランダムプライマーにアダプター配列を付与するcell barcoding技術の開発を行った。アダプター配列をつけたランダムプライマーをつかった分子生物学的な手法はこれまでも報告されているが、集団細胞を対象としているので、効率は不明である。効率が低いと、1細胞では定量性能の側面から大きな問題となる。まず様々な長さのアダプター配列をN6ランダム配列の5'側に付与して、逆転写反応を行った。結果としては、アダプター配列の長さが35塩基程度以上だと、逆転写効率が1/7-1/8程度以下になることが明らかになった。逆に、アダプター配列を17塩基程度以下にすると、効率が回復してくることがわかった(図7)。この結果から、ランダムプライマーでのcell barcodingに使用する配列は少なくとも17塩基長以下で選択する必要が有ることを示唆している。

次にcell barcodingに使用するcell barcode配列の計算を行った。Cell barcodeとして正確に認識・割り振るためには、お互いがユニークであり、シーケンスエラーなどを補正出来ることが望まれる。そこで、團野宏樹研究員と連携し、シーケンスレーベンシュタイン距離(編集距離5)の12塩基の配列を計算した。これは、塩基欠失や塩基置換が2個あってもどちらの配列由来か判定できる。結果、226個の配列が見つかった。一方で、塩基組成に偏りがあることが観察された(図8)。この偏りは、Illumina sequencerで解析する際に問題になる可能性があった。そのため、226種類から96種類選択して塩基組成の偏りを減らすことに成功した(図9)。またハミング距離(編集距離5)で、なおかつ6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基でも計算し配列を求めた。結果、それぞれ8種類、20種類、48種類、125種類、363種類、953種類、2857種類の配列が得られた。塩基配列としても、シーケンスレーベンシュタイン距離の場合に比べて、塩基偏りは緩やかであった(図10)。9塩基あれば、96 well plate 1枚分で、11塩基あれば384 well x 2枚分でCell barcodingが可能になると予想される。上述の開発した1細胞RNA-seqの反応と組み合わせ現在single-cell RNA-seqの開発・解析を進めている。また解析環境も必要となるので、ハイスループットsingle-cell RNA-seq法であるDrop-seqの解析パイプラインを市販PCの

Mac pro にインストールした。別プロジェクトで申請者が開発したハイスループット single-cell RNA-seq 法である Quartz-Seq2 のデモデータを解析できることを確かめた。また Cell-barcoding の認識を別部位に変えることで、同パイプラインを上記のランダムプライマーを用いた cell barcoding 技術によるハイスループット single-cell RNA-seq 法に流用することが可能となった。

[まとめ]

本課題では、1 細胞マルチオミックス技術の開発を目指した。課題推進中に、提案した方法の一部が達成されてしまったが、途中で方向転換して、新生 RNA を含む RNA 分子を捉えられる single-cell RNA-seq 法を確立することに成功した。また、cell barcoding 技術をランダムプライマーに持ち込むことを検討し、高出力化への道筋をつけた。

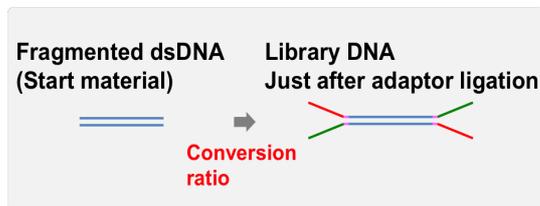


図 1. Y 型アダプターライゲーションによるシーケンスライブラリ作製方法

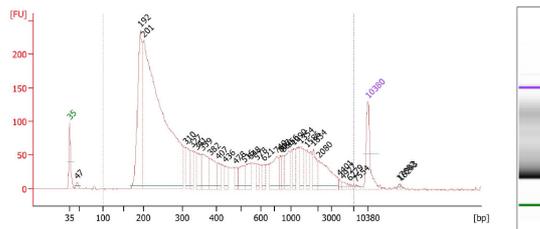


図 2. Single-cell ATAC-seq のライブラリ DNA のサイズ長分布

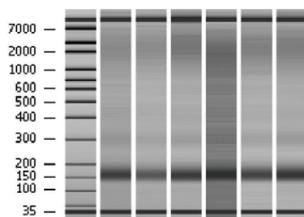


図 3. dsDNase 処理した核の nucleosome pattern

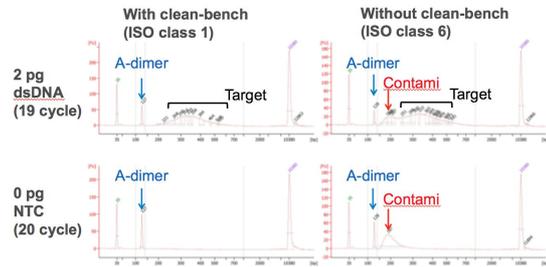


図 4. 微量 DNA からのライブラリ作製

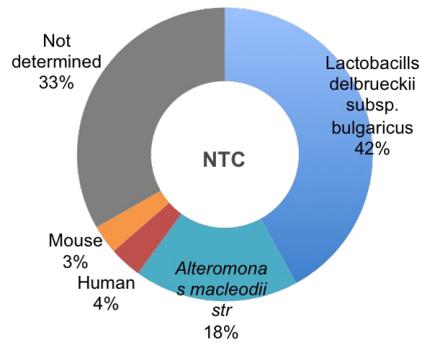


図 5. 実験環境中の浮遊 DNA のメタゲノム解析

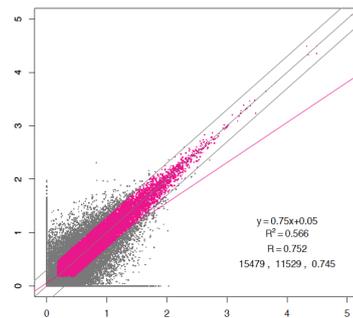


図 6. 開発方法を用いた 10 pg total RNA からの 2 回の発現定量の再現性

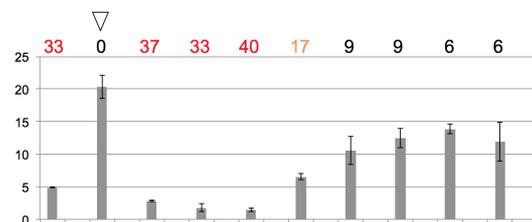


図 7. アダプター配列の長さによる逆転写効率への影響

Y 軸は、qPCR score。棒グラフの上部にある数字が、アダプター配列の長さ。

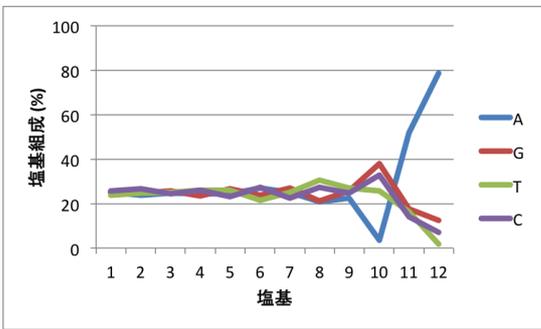


図 8. シーケンスレーベンシュタイン距離セルバーコード配列 (12 塩基) の塩基組成

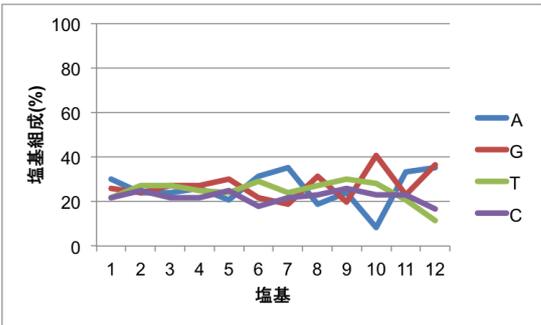


図 9. 補正後のシーケンスレーベンシュタイン距離セルバーコード配列 (12 塩基) の塩基組成

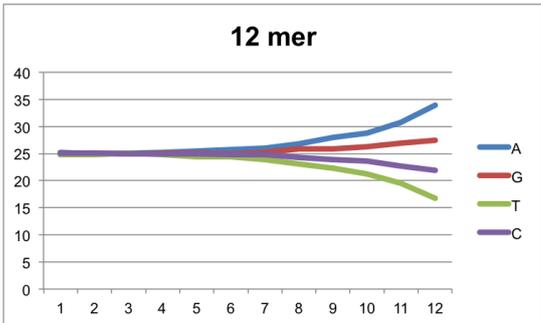


図 10. ハミング距離セルバーコード配列 (12 塩基) の塩基組成

5. 主な発表論文等  
(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 2 件)

[1] 笹川洋平, 團野宏樹, 海老澤昌史, 林哲太郎, 二階堂愛、セルソーターを用いたハイスループット 1 細胞 RNA-seq 法 Quartz-Seq2 の開発、第 39 回日本分子生物学会年会、2016 年 12 月 2 日、パシフィコ横浜 (神奈川県)

[2] Yohei SASAGAWA, Tetsutaro Hayashi,

Itoshi Nikaido、non-poly-A RNA も検出できる 1 細胞 RNA-seq 方の開発と比較、NGS 現場の会 第四回研究会、2015 年 7 月 1 日、つくば国際会議場 (茨城県)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況 (計 0 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

6. 研究組織

(1) 研究代表者

笹川洋平 (Sasagawa Yohei)  
国立研究開発法人理化学研究所・情報基盤センター・上級センター研究員  
研究者番号：10404344

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

團野宏樹 (Danno Hiroki)  
国立研究開発法人理化学研究所・情報基盤センター・センター研究員  
研究者番号：20610439

(4) 研究協力者

( )