

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：82508

研究種目：基盤研究(C)

研究期間：2010～2013

課題番号：22510217

研究課題名(和文)核移行・核外移行に関わるシグナル配列の高精度統合予測システムの開発と利用

研究課題名(英文)Development of computational system to accurately predict nuclear localization signals and nuclear export signals in proteins

研究代表者

小杉 俊一 (Kosugi, Shunichi)

公益財団法人かずさDNA研究所・植物ゲノム研究部・プロジェクト研究員

研究者番号：30365457

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：本研究では、核移行シグナル予測ツールの植物対応版(cNLS Mapper II)の開発、および核外移行シグナル予測ツール(NESmapper)の開発を行った。NES予測ツールの開発では、4つのNESクラスについて作成したNES活性プロファイルを基にNESmapperを開発した。テストデータを用いたNES予測テストでは、ネガティブデータからのNES予測を10%以下に抑え、ポジティブデータの8割以上を予測することに成功した。開発したNES予測プログラムの予測精度は、既存のNES予測ツールの予測精度を上回るものであり、最も有用なNES予測ツールとして提供できることが期待される。

研究成果の概要(英文)：In this study, I developed a plant-optimized version of cNLS Mapper, a prediction tool of nuclear localization signals, by creating an activity-based profile of a plant-specific NLS class. I also developed a computational tool (NESmapper) for nuclear export signal (NES) prediction, by using newly created activity-based profiles for all the NES subclasses. An evaluation test for NES prediction with test data sets showed that NESmapper predicted over 80% of positive NESs with 10% of false positive rate. Comparison of the prediction performance of NESmapper with those of other available tools (i.e., NESsential, NetMES, and Wregex) demonstrated that NESmapper outperforms other tools. NESmapper and also its combined version with cNLS Mapper will be provided for the scientific community as a valuable tool to predict functional NESs and NLSs.

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・システムゲノム科学

キーワード：ゲノム プロテオーム バイオインフォマティクス 解析技術 核移行シグナル 核外移行シグナル
モチーフ 予測

1. 研究開始当初の背景

転写因子を含めた核で機能するタンパク質の多くは、自身の持つ核移行シグナル(NLS)の働きで核に移行する。さらに、それらのタンパク質の多くは、核から細胞質への移行を指令する核外移行シグナル(NES)を持っている。NLSとNESは、互いの働きを拮抗させながら、タンパク質の核—細胞質間の移行制御を行い、核タンパク質の機能を制御している。NLSおよびNESを精度高く予測することができれば、タンパク質の核移行制御機構を類推でき、効率的に解析を進めることができるようになる。

(1) 核移行シグナルの予測 importin α によって認識される classical NLS は、NLSの中で最も主要なクラスを構成している。塩基性アミノ酸クラスターからなる NLS コンセンサス配列が古くから知られるが、配列の厳密性が低いため精度高く NLS を予測することは困難である。申請者は、importin α に特異的に結合する 6 クラスの NLS を同定した(文献 1)。さらに、各 NLS クラスについて体系的アミノ酸置換解析を行い、NLS 配列内アミノ酸の核移行活性影響度を示すプロファイルを作製した(文献 2, 3)。このプロファイルと NLS 内アミノ酸の相加性特性を利用したアルゴリズムを用いて、NLS 活性を計算するコンピュータプログラム(cNLS Mapper, <http://nls-mapper.iab.keio.ac.jp/>)を開発した(文献 2)。cNLS Mapper は、90%以上の NLS 予測精度を示し、既存の NLS 予測ツール(Psort II, Predict NLS)の予測精度を特異性、感度共に大きく上回っていた(文献 2)。この予測法の基本原理であるアミノ酸相加性特性を利用してデザインされたペプチドが、importin α に対して極めて高い結合親和性を持ち、核移行阻害剤として機能することは、この原理が NLS に対して正しく働いていることを示すものである(文献 3)。cNLS Mapper は、出芽酵母を用いて作製した NLS プロファイルを基にして作成されているが、importin α を介したタンパク質核移行経路は真核生物において高度に保存されているため、酵母以外の生物種における予測にも適用できると考えられる。しかし、植物特異的に機能する NLS クラス(class 5 NLS)の存在や、他の NLS クラスにおいてもアミノ酸特異性が酵母と植物間で異なる例が多く観察されている(文献 1)。このことから、植物の NLS を精度高く予測するためには、植物を用いたアミノ酸置換解析によってプロファイルを作製し、植物に対応した NLS 予測プログラムを開発する必要がある。

(2) 核外移行シグナルの予測 NES は、importin β ファミリーのメンバーである exportin (CRM1)によって認識される。NES の古典的コンセンサス配列(ϕ -X_{2,3}- ϕ -X_{2,3}- ϕ -X- ϕ , ϕ : L, I, V, M, or F)は、他の多くのモチーフコンセンサス配列同様厳密性が低く、シロイヌナズナのほぼ

全タンパク質にこのコンセンサス配列がヒットしてしまう。最近定義されたより厳密度を増した NES コンセンサス配列(文献 4)においても予測精度の上昇はあまり見られない。文献から収集した NES 配列を基に、コンピュータ学習法によって NES を予測する試みがなされているが(文献 5)さほど高い予測精度は得られていない。

NES コンセンサス配列の極めて低い特異性は、その疎水性アミノ酸に富んだ配列が、タンパク質の内部に折り畳まれた領域や膜局在領域などの疎水性アミノ酸に富んだ領域に(偽 NES 配列として)多く含まれることによる。NES コンセンサス配列の疎水性アミノ酸および任意アミノ酸位置では、正や負に様々なレベルで NES 活性に作用するアミノ酸が存在する(文献 4, 5)。このようなアミノ酸による影響度は、NLS 同様、体系的アミノ酸置換解析に基づいて作製されたプロファイルによって明らかにされるはずである。したがって、NES の高精度予測プログラムは、NLS の場合と同じ手法によって開発できると考えられる。

2. 研究の目的

NLS、NES 配列の正確な予測は困難であるため、本研究では次の 4 つの段階からなる、高精度 NLS/NES 予測システムの開発を行う。① 最近応募者が開発した高精度 NLS 予測プログラム cNLS Mapper の植物対応版の開発、② 新規の NES 予測プログラムの開発、③ NLS/NES 統合予測システムの開発。④ さらに、開発した予測システムを用いて、シロイヌナズナから NLS、NES によって移行制御を受けるタンパク質の予測、同定を行う。本研究の最大の目的は、応募者が独自に開発した高精度 NLS 予測法(モチーフ内アミノ酸の活性影響度を示すプロファイルを基に、アミノ酸相加性特性を利用してモチーフ活性を計算する手法)を NES の予測に適用することによって、今まで成し得なかった高精度の NES 予測プログラムを開発することにある。さらに、NLS 予測プログラム(cNLS Mapper)やタンパク質のトポロジーや構造特性に関わる予測プログラムを組み合わせることによって、NLS/NES 統合予測システムを開発する。この予測システムは、植物研究分野だけでなく真核生物全般のタンパク質細胞内移行制御の解析に有用なツールを提供するものと期待される。

3. 研究の方法

(1) 植物 NLS に対応した高精度 NLS 予測プログラムの開発
高精度 NLS 予測プログラム(cNLS Mapper)は、酵母に至適化された NLS プロファイルを基に開発されたため(文献 1)植物由来の NLS をより精度高く予測するためには、植物に至適化された NLS プロファイルを作製する必要がある。

(a) プロファイルの作製方法：プロファイルの作製は、文献 2、3 に従って行う。まず各クラスの NLS の代表的な鑄型配列 (NLS 配列をコードするオリゴヌクレオチド) を用いて、配列内の各アミノ酸位置についてアミノ酸置換解析を行う。変異オリゴヌクレオチドをクローニングする際に、できるだけ多くの種類のアミノ酸置換体をシークエンスによって選択する。クローニングに用いるベクターは、pTUE-GFP3 ベクターを用いる。オリゴヌクレオチドを挿入した本ベクターを細胞に導入すると、GUS-GFP-NLS 融合タンパク質が発現され、NLS に核移行活性があると GFP の蛍光は核に、活性が無いと細胞質に検出される。このときの核移行度 (核局在度とその細胞の割合) を基に NLS の活性を 10 段階に定量化する。

(b) 予測プログラムの作成：pTUE-GFP3 は、tTA 転写因子を共発現させることによって、酵母、植物、ほ乳動物で共通して使用できる (文献 1)。このため、酵母のプロファイルの作製に用いた各変異 NLS を挿入した pTUE-GFP3 プラスミドは、そのまま植物細胞を用いた核移行アッセイに用いることができる。タバコ培養細胞を用いたアッセイによって、植物に至適化された class 1/2 NLS などのプロファイルはすでに完成している。あとは植物特異的 NLS である class 5 NLS などのプロファイルを作成させることで、植物に対応した NLS 予測プログラムを作成することが可能な状況にある。作成したプログラムは、class 5 NLS を含むテスト配列を用いて予測精度の評価を行い、cNLS Mapper II として公開する。

(2) 高精度 NES 予測プログラムの開発

NES 予測プログラムの開発は、NLS 予測プログラムの場合と同様、体系的アミノ酸置換解析に基づいたプロファイリングによる方法で行う。申請者による予備実験から、NES 活性は保存された疎水性アミノ酸だけでなく、非保存位置においても、アミノ酸の種類によって活性影響度が大きく異なることが観察されている。このことから、プロファイルに基づいた予測法が NES に対しても有効であることが示唆される。本研究が対象とする CRM1 依存的 NES は、疎水性アミノ酸の配列パターンによって計 6 クラスに分類される (文献 4)。各 NES クラスのアミノ酸置換解析は、pTUE-GFP2 ベクター (文献 2) を用いて行う。このベクターも酵母や植物での使用に対応した発現ベクターであるが、レポーターとして GFP のみを含む。細胞内で発現される GFP-NES 融合タンパク質は、NES が核外移行活性を持つ場合は細胞質に、活性を持たない場合は細胞全体に観察されるため、この局在度を指標として NES の活性を定量化する。作成した予測プログラムは、各 NES クラスのテスト配列を用いて、植物だけでなく酵母や哺乳動物の NES に対して、どの程度の予測精度を達成しうるかを検証

する。

(3) NLS/NES 高精度統合予測システムの開発

NLS、NES などのモチーフの多くは、タンパク質の非構造領域に存在する。また、膜タンパク質は、NLS や NES が存在してもそれらの活性の影響を受けないことが申請者の予備実験で明らかとなっている。さらに、疎水性アミノ酸に富んだ膜局在化領域は、多くの偽 NES 配列を含むことがわかっている。このことから、タンパク質の非構造領域や膜局在化ドメインを予測する既存ツールを組み合わせて予測を行うことにより、より精度高く NLS/NES を予測できると考えられる。これらの既存ツールを cNLS Mapper II と NES 予測ツールに組み込み、統合させることによって、NLS/NES 高精度予測システムを開発する。

(4) 核への出入りによって制御を受けるタンパク質の予測と同定

植物で核と細胞質間を行き来して機能制御を受けるタンパク質を同定するために、開発した統合予測システムを用いて、シロイヌナズナの全タンパク質配列から NLS および NES を予測し、両モチーフを含むタンパク質を抽出する。本研究では、この中から機能情報および細胞内局在性情報を基に 20 から 30 タンパク質を選択し、実験検証により核細胞質間を行き来して機能制御を受けるタンパク質を同定する。実験検証は、exportin 阻害剤である leptomycin B (LMB) および importin α 阻害剤である bimax1 (文献 3) を用いて、試験タンパク質の細胞内局在性変化を観察することによって評価する。

【文献】(1) Kosugi S. et al. *J. Biol. Chem.* **284**, 478-485 (2009)

(2) Kosugi S., Hasebe M., Tomita M., and Yanagawa H. *PNAS* **106**, 10171-10176 (2009)

(3) Kosugi S. et al. *Chem. Biol.* **15**, 940-949 (2008)

(4) Kosugi S., Hasebe M., Tomita M., and Yanagawa H. *Traffic* **9**, 2053-2062 (2008)

(5) la Cour T. et al. *Protein Eng. Des. Sel.* **17**, 527-536 (2004)

4. 研究成果

(1) 植物 NLS に対応した高精度 NLS 予測プログラムの開発

植物特異的な class 5 NLS のプロファイルを作製し、このプロファイルを取り入れた植物対応版 cNLS Mapper II を作成した。class 5 テスト NLS 配列を用いて予測精度検証試験を行った結果、class 5 NLS プロファイルを取り入れていない従来の cNLS Mapper では、37%の class 5 NLS しか予測出来なかったのに対して、cNLS Mapper II では、100%の class 5 NLS を予測出来た【表 1】。

【表 1】cNLS Mapper II の class 5 NLSs の予測精度

Hit number for 27 test class 5 NLSs	
cNLS Mapper	cNLS Mapper II
10 (37%)	27 (100%)

NLS 活性の実測スコアと予測スコアの差が 2 以内の場合をヒットとした

(2) 高精度 NES 予測プログラム (NESmapper) の開発

4 つの NES サブクラス (class-1b, class-1c, class-2, および class-1a と class-3 を融合させた class-1a/3) について、各 NES クラスの代表的 NES 配列を鋳型として、各位置のアミノ酸を他の 8-19 アミノ酸にランダムに置換した変異体を作製した。得られた変異体のプラスミドクローンをマウス培養細胞に導入し、一過的に発現された GFP-NES 融合タンパク質の核外移行度を 10 段階に区分して表記し、鋳型 NES 配列の各位置での単一アミノ酸置換により変化する核外移行活性を表す NES プロファイルを各 NES クラスについて作成した。Class-1a/3 NES のプロファイルを【図 1】に示す。

	Φ0	Φ1	Φ2	Φ3	Φ4									
	P ^a	S	S	L	A	E	K	L	A	G	L	D	I	N
L	9	9	8	8	7	8	7	8	8	8	8	8	9	8
I	11	9	8	8	7	7	8	8	8	8	7	8	8	8
V	11	9	8	8	7	8	8	10	8	8	8	8	7	7
M	10	9	8	8	8	8	8	8	8	7	8	8	7	7
C	8	8	8	6	7	8	8	8	8	8	6	8	4	8
F	8	8	8	8	8	7	7	6	6	6	5	6	3	7
Y	8	8	8	3	8	8	5	8	8	8	1	7	0	7
W	8	8	8	4	7	1	8	8	8	2	8	1	3	8
T	8	8	5	8	9	8	3	6	8	3	9	2	8	8
P	8	6	2	5	5	4	-5	2	2	-5	-2	-5	0	8
A	8	8	6	8	8	8	2	8	8	2	6	1	8	8
G	8	8	0	8	8	8	-1	9	8	-2	-1	8	8	8
S	6	8	8	1	8	9	8	-1	8	8	-2	10	-1	8
H	8	8	0	8	7	-1	8	8	-2	8	-1	7	8	8
R	8	8	0	8	8	8	-1	8	6	-2	8	-1	8	8
K	8	8	0	8	8	8	-1	8	8	-2	10	-1	8	8
Q	8	8	2	8	8	7	-1	8	8	8	8	-1	8	8
N	8	8	2	8	8	8	-1	5	8	8	3	-1	8	8
D	9	9	8	8	8	8	8	8	8	8	8	-1	10	8
E	9	9	0	8	8	8	-1	8	8	8	8	-1	8	8

【図 1】Class-1a/3 NES プロファイル

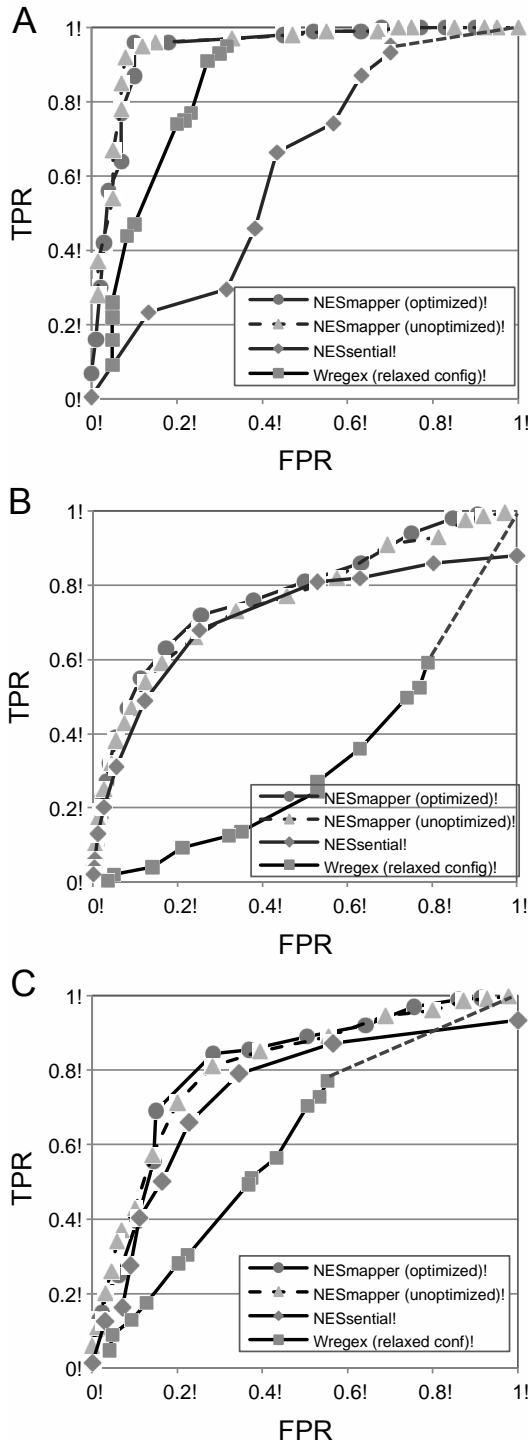
最上行に示される鋳型 NES 配列の各位置のアミノ酸を左列に示されるアミノ酸で置換した配列の核外移行活性スコアを-5 から 11 の数値で示す。鋳型の標準活性スコアは 8 である。標準スコアより大きく変化した箇所をハイライト表示した。Φ1-Φ4 は、保存疎水性アミノ酸位置。

各プロファイルを取り入れて NES の核外移行活性を計算・予測するコンピュータプログラム NESmapper を作成した。NESmapper の予測精度を向上させるため、2 つの改良を加えた。1 つは、NES の上下流領域のアミノ酸組成を計測し、NES 予測に取

り入れた。これは、ポジティブおよびネガティブ NES データについて、25 アミノ酸 N 末上流および C 末下流領域を比較した結果、ポジティブ NES の N 末上流では極性アミノ酸の比率が、C 末下流領域ではアミノ酸正味荷電がネガティブ NES のものに比較して有為に高いことが観察されたことによる。2 つ目は、NES プロファイルをコンピュータトレーニング手法を用いて最適化した。実験によって計測できなかったプロファイル内のブランク箇所は、他のプロファイルデータやアミノ酸類似性を基に類推して補填する必要があった。これらの類推値は不正確である場合があるため、ポジティブおよびネガティブ NESs のトレーニングデータセットを用いて、NESmapper によるデータセットの予測値と実測値が近くなるように、プロファイルのスコア値を微調整し、NES プロファイルの最適化を行った。このトレーニング手法では、自作したコンピュータプログラムを用いて、トレーニングデータの予測精度が上がるまで各 NES プロファイルの中の各スコアを上下させ、この一連の操作を数百回繰り返すことでプロファイルの最適化を行った。また、膜タンパク質等に見られる疎水性アミノ酸に富んだ領域で多くの擬陽性 NES を検出してしまいう問題については、当初は疎水性領域を予測する外部プログラムの導入を計画していたが、NES プロファイルや上記改良法に加え、NES の非保存領域中の疎水性アミノ酸数の割合に制限を加える改良を加えることで解決することが出来た。

(3) NESmapper の NES 予測精度の評価

NESmapper の NES 予測精度を検証するために、ポジティブ およびネガティブ NES のテストデータセットをそれぞれ 2 種類用意した。ポジティブデータセットは、NES データベースから取得した 205 の LMB 感受性 NESs (ValidNES-dataset) 人工ペプチドライブラリーから取得し、NES 活性を測定した 163 の NESs (artificial dataset)。ネガティブデータセットは、LMB で細胞内局在性の変化が観察されなかった分裂酵母の 60 の細胞質局在タンパク質 (231 の NES コンセンサス配列を含む) (Sp-dataset) 人工ペプチドライブラリーから取得し、NES コンセンサス配列を含むが NES 活性の無いことを確認した 92 の NESs (artificial dataset)。これらのデータセットについての統合した結果から、NESmapper はネガティブデータからの NES 予測を 10%以下に抑え、ポジティブデータの 8 割以上を予測することができた。既存の NES 予測ツール (NetNES, NESsential, Wregex) の予測精度と比較しても、いずれのデータセットにおいても NESmapper の予測精度が上回っていた。最近発表された 2 つのツールについての予測精度比較を Receiver Operator Characteristic (ROC) 解析によって比較した結果を【図 2】に示す。



【図2】NESmapper、NESsential、WregexのNES予測精度についてのROC解析 (A) Artificial NES datasetsを用いて作成したROC曲線 (B) ValidNES/Sp-datasetsを用いて作成したROC曲線 (C) Artificial datasetsとValidNES/Sp-datasetsを用いて作成したROC曲線 TPR: true positive rate, FPR: false positive rate

精度の指標となる the area under the curve (AUC)の値は、NESsential(0.75)やWregex(0.63)と比較してNESmapperのもの(最適化プロファイル: 0.81、最適化プロファイ

ル: 0.80)が最も高かった。これらの結果から、開発したNESmapperは、既存のNES予測ツールより高い精度を持つ有用なツールであることが示される。

核外移行制御を受けるタンパク質の多くは、NESの他にNLSを併せ持つことが知られる。今後、NESmapperとcNLS Mapper IIを統合し、NESとNLSを同時に予測することにより、利便性を上げるだけでなく、NESとNLSの予測精度をさらに向上させることが可能になると期待される。

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10件)

Masafumi Yagi*, Shunichi Kosugi*, Hideki Hirakawa et al. (他24名)

Sequence Analysis of the Genome of Carnation (*Dianthus caryophyllus* L.). DNA Research、査読有、2013、Epub ahead of print

DOI: 10.1093/dnares/dst053

* Co-first authors

Shunichi Kosugi, Satoshi Natsume, Kentaro Yoshida, Daniel MacLean, Liliana Cano, Sophien Kamoun, and Ryohei Terauchi

PLoS One、査読有、8、2013、e75402
DOI: 10.1371/journal.pone.0075402

Hiroki Takagi, Akira Abe, Kentaro Yoshida, Shunichi Kosugi, Satoshi Natsume, Chikako Mitsuoka, Aiko Uemura, Hiroe Utsushi, Muluneh Tamiru, Shohei Takuno, Hideki Innan, Liliana M. Cano, Sophien Kamoun, and Ryohei Terauchi

Plant Journal、査読有、74、2013、174-183
DOI: 10.1371/journal.pone.0083720

Rym Fekih, Hiroki Takagi, Muluneh Tamiru, Akira Abe, Satoshi Natsume, Hiroko Yaegashi, Shailendra Sharma, Shiveta Sharma, Hiroyuki Kanzaki, Hideo Matsumura, Hiromasa Saito, Chikako Mitsuoka, Hiroe Utsushi, Aiko Uemura, Eiko Kanzaki, Shunichi Kosugi, Kentaro Yoshida, Liliana Cano, Sophien Kamoun, and Ryohei Terauchi

PLoS One、査読有、8、2013、e68529
DOI: 10.1371/journal.pone.0068529

Jeffrey A. Fawcett, Tomoyuki Kado, Eriko Sasaki, Shohei Takuno, Kentaro Yoshida, Ryuichi P. Sugino, Shunichi Kosugi, Satoshi Natsume, Chikako Mitsuoka, Aiko Uemura, Hiroki Takagi, Akira Abe, Takashige Ishii, Ryohei Terauchi, and Hideki Innan

PLoS One、査読有、8、2013、e83720

DOI: 10.1371/journal.pone.0083720
Akira Abe*, Shunichi Kosugi*,
Kentaro Yoshida, Satoshi Natsume,
Hiroki Takagi, Hiroyuki Kanzaki,
Hideo Matsumura, Kakoto Yoshida,
Chikako Mitsuoka, Muluneh Tamiru,
Hideki Innan, Liliana Cano, Sophien
Kamoun, and Ryohei Terauchi
Nature Biotechnology、査読有、30、2012、
174-178
DOI: 10.1038/nbt.2095
* Co-first authors
Kosuke Okazaki1, Naomi Nakayama,
Yuko Nariai, Hiroaki Kato, Kentaro
Nakayama, Kohji Miyazaki, Riruke
Maruyama, Shunichi Kosugi, Takeshi
Urano, and Gyosuke Sakashita
Carcinogenesis、査読有、33、2012、
1854-1862
DOI: 10.1093/carcin/bgs193
小杉俊二、柳川弘志
タンパク質モチーフ活性の予測とシグ
ナル伝達解析
生化学、査読有、82、2010、37-40
http://www.jbsoc.or.jp/back_no/82-7

〔学会発表〕(計2件)

小杉 俊一

Coval: Improving alignment quality
and variant calling accuracy for
next-generation sequencing data
日本バイオインフォマティクス学会
2013年年会、2013年10月29-31日、タ
ワーホール船堀

小杉 俊二、平川 英樹、田畑 哲之
ゲノムアセンブリのギャップを高精度で
埋めるツール GMcloser の開発
NGS 現場の会 第3回研究会、2013年9
月4-5日、神戸国際会議場

〔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

小杉 俊一 (KOSUGI, Shunichi)
公益財団法人かずさDNA研究所・植物ゲ
ノム研究部・プロジェクト研究員
研究者番号：30365457

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし