

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 22 日現在

機関番号：82657

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26830129

研究課題名(和文)トランスクリプトームを基盤とする統合遺伝子検索システムの構築

研究課題名(英文)Transcript-oriented search engine for genes and transcripts

研究代表者

内藤 雄樹(Naito, Yuki)

大学共同利用機関法人情報・システム研究機構(新領域融合研究センター及びライフサイ・ライフサイエンス統合データベースセンター・特任助教)

研究者番号：60451829

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：統合遺伝子検索システム構築のため、塩基配列を効率よく検索することのできるGGGenome (<http://GGGenome.dbcls.jp/>) というソフトウェアの開発を進めた。RefSeqに収録された転写産物だけでなく、塩基配列が公開されている200種以上の生物のゲノムや国際塩基配列データベース上の全塩基配列を検索できるようにするとともに、利用者が簡単に検索できるウェブインターフェースを開発した。開発したソフトウェアは、ゲノム編集のためのガイドRNA設計や、核酸医薬品のオフターゲット効果の予測にも有効であった。本成果は、ウェブツールとして無償で公開し、誰でも自由に活用できるようにした。

研究成果の概要(英文)：I have developed GGGenome (<http://GGGenome.dbcls.jp/>), a fast and sensitive search engine for nucleotide sequences. GGGenome provides simple and easy-to-use web interface for searching not only RefSeq transcripts, but also genomic sequences for more than 200 organisms and the entire sequences of the International Nucleotide Sequence Database. The software is widely applicable for off-target searches of CRISPR/Cas9 guide RNAs and oligonucleotide therapeutics. All services of the web servers developed in this work are freely available to all users.

研究分野：分子生物学、生物情報科学

キーワード：遺伝子検索システム トランスクリプトーム 塩基配列検索 GGRNA GGGenome CRISPRdirect

1. 研究開始当初の背景

データベースから遺伝子を検索するという作業は、多くの生命科学系あるいは医学薬学系の研究者にとり日常的なものである。しかし、既存のデータベースを利用して目的の遺伝子の情報をすばやく探しだすことは必ずしも容易でない。たとえば、検索の目的やキーワードの種類に応じてさまざまなデータベースを使い分ける必要があり、利用者にとって煩雑であるばかりでなく、その検索に適したデータベースの在処を把握していなければ効率的に情報を得ることができない。

また、GenBankをはじめとする既存の塩基配列データベースは、ひとつの遺伝子に対し複数のエントリが存在するなど冗長であることが多く、単純に遺伝子名で検索する場合でさえも多数のエントリがヒットし目的の情報になかなかとり着けない場合が多い。さらに、塩基配列やアミノ酸配列の検索において広範に利用されている BLAST 検索のウェブサーバは、検索に十秒から数十秒もかかるなど、利用者が不便を感じるが多かった。

そこで研究代表者は、あらゆるキーワードや配列情報をひとつの検索窓に入力するだけで、迅速に遺伝子や転写産物を探せるようなウェブサーバを構築したいと考えた。研究代表者は、そのプロトタイプとして、NCBI RefSeq で提供されている生物種のうちモデル生物 13 種を全文検索可能な「統合遺伝子検索システム GGRNA」を 2012 年より公開している。

2. 研究の目的

あらゆるキーワードや配列情報から遺伝子をすばやく検索し、その結果をわかりやすく提示することのできるような統合遺伝子検索エンジンを開発して広く公開する。そのため、とくに塩基配列を効率よく検索することのできるシステムを開発する。これにより、いままで煩雑であった遺伝子を検索するという作業が、簡単かつ短時間でできるようになり、広く生命科学分野あるいは医学薬学分野の研究に貢献できると期待される。

3. 研究の方法

研究代表者が開発を進めてきた統合遺伝子検索 GGRNA (<http://GGRNA.dbcls.jp/>) の機能をもとに、塩基配列を効率よく検索することのできるシステムを開発する。塩基配列の検索に注力して開発を行う理由は、塩基配列の検索は一般のテキストを検索する場合と異なり、文字列が完全に一致せず少数の不一致や挿入欠失を含む場合であっても見落とさず検索したいという需要があり、とくに短い塩基配列でそれを効率よく実行できるサービスが十分に提供されていないた

めである。

なお、本研究の成果はウェブツールとして無償で公開し、すべての研究者が自由に活用できるようにする。

4. 研究成果

本研究では、研究代表者がこれまでに開発してきた統合遺伝子検索システム GGRNA (<http://GGRNA.dbcls.jp/>) を改良するとともに、塩基配列を効率よく検索することのできるソフトウェア GGGenome (<http://GGGenome.dbcls.jp/>) の開発を行った。GGGenome は、一般的な塩基配列検索ツールでは検索の困難な 20 塩基未満の配列でもすばやく検索することができ、配列に N, R, Y などあいまいな塩基を含む場合や、配列全体の 4 分の 1 程度まで mismatches や挿入欠失を含む配列であっても、見落とさず検索することができる。RefSeq に収録された転写産物だけでなく、塩基配列が公開されている 200 種以上の生物のゲノムや、DDBJ/ENA/GenBank 国際塩基配列データベース上の全塩基配列を検索できるようにした。また、外部のソフトウェアやデータベースとの連携を容易にする REST API を整備した。

さらに、GGGenome による塩基配列の検索が役立つと考えられる生物学の課題を検討した。その一例は、CRISPR/Cas9 ゲノム編集法におけるガイド RNA の設計である。CRISPR/Cas9 法では、標的とする部位以外のゲノムを誤って認識してしまうオフターゲット効果を防ぐために、ゲノム全体の塩基配列を検索し、ほかの領域とは配列類似性の低い領域に対しガイド RNA を設計する必要がある。このような検索は従来の塩基配列検索ツールでは困難であったが、GGGenome を用いることにより容易となった。前述の GGGenome の REST API を活用することにより、CRISPR/Cas9 法のためのガイド RNA 設計ソフトウェア CRISPRdirect (<http://crispr.dbcls.jp/>) を開発して公開した。さらに、核酸医薬品のオフターゲット効果の予測にも、GGGenome による検索が有効であることがわかった。

本研究により開発したソフトウェアはウェブツールとして無償で公開し、すべての利用者が自由に活用できるようにした。以下に、本研究で開発を行った GGRNA、GGGenome、CRISPRdirect の詳細を述べる。なお、以下の内容は既に文献 (5. 主な発表論文等 [雑誌論文]) に報告した。

(1) GGRNA (<http://GGRNA.dbcls.jp/>)
[キーワードや配列から遺伝子を検索]

このツールでできること

- ・ 遺伝子や転写産物をあらゆる語句から検索
- ・ 塩基配列やアミノ酸配列の完全一致検索

このツールの特徴

GGRNA (ググルナ) は、遺伝子や転写産物を Google のようにすばやく検索できるウェブツールである。検索キーワードとして、遺伝子名やアクセッション番号など各種の ID をはじめ、遺伝子の機能やタンパク質のドメイン名、さらには、塩基配列やアミノ酸配列など、あらゆる語句をひとつの検索窓に入力するだけで、RefSeq に登録された転写産物をすばやく探し出すことができる。とくに、塩基配列やアミノ酸配列の検索においては、一般的な配列類似性検索ツールでは検索の困難な 10 塩基あるいは 4 アミノ酸ほどの短い配列でも高速な検索が可能である。

ツールの名前は、GooGle ライクな転写産物 (RNA) の検索エンジンを意味するが、ツール公開後に「ググルナ」という呼称がしだいに定着していった。Google のように強力ですべて簡単に使える検索エンジンをめざし、遺伝子の検索においてはググルなくてもこのツールで効率的に情報を得られるようにしたいという目標をもち開発を進めている。

(2) GGGenome (<http://GGGenome.dbcls.jp/>) [高速な塩基配列の検索ツール]

このツールでできること

- ・高速で漏れのない塩基配列の検索
- ・ミスマッチや挿入欠失を許容した検索
- ・とくに短い配列の検索に強い

このツールの特徴

GGGenome (ゲゲゲノム) は、ゲノムなどの塩基配列を高速に検索することができるウェブツールである。ツールの名前は、前述した GGRNA の「RNA」を「Genome」に置き換え GGGenome とした。当初はこの名前が示すとおり、代表的なモデル生物のゲノムを検索できるツールを想定していたが、ゲノムだけでなく RefSeq に収録された転写産物、さらには DDBJ / ENA / GenBank 国際塩基配列データベースに登録されたすべての塩基配列を効率よく検索することのできるシステムを開発した。塩基配列の検索プログラムは、ほかにも BLAST や BLAT が有名だが、これらのツールでは検索の困難な 20 塩基未満の配列でも、GGGenome によりすばやく検索することができる。このような短い配列の検索は、PCR のプライマー、プローブ、siRNA、タンパク質などの結合サイトを調べるのにも有用と思われる。GGGenome では、配列に N, R, Y などあいまいな塩基を含む場合や、配列全体の 4 分の 1 程度までミスマッチや挿入欠失を含む配列であっても、漏れのない検索が可能である。このことも、ほかのツールにはない GGGenome の大きな強みである。一方、1,000 塩基を超える長い配列の検索や、スプライシング後の cDNA 配列をエキソン・イントロン構造を考慮してゲノムにマッピングしたい場合は、GGGenome よりも BLAT が適しているのもそちらを利用する必要がある。

(3) CRISPRdirect (<http://crispr.dbcls.jp/>) [CRISPR/Cas9 ゲノム編集法のためのガイド RNA 設計ツール]

このツールでできること

- ・特異性の高いガイド RNA の設計
- ・オフターゲット部位の予測

CRISPR/Cas9 システム

近年、ゲノム DNA の編集を行うための手法として、CRISPR/Cas9 システムを用いたゲノム編集法が注目されている。この手法は、Cas9 とよばれる RNA 誘導型エンドヌクレアーゼおよび標的認識を担うガイド RNA という 2 つの因子の相互作用により、ガイド配列と相補的なゲノム DNA の切断を引き起こす手法である。塩基配列特異的な DNA の切断を利用して、ゲノム上の任意の場所で遺伝子の機能破壊やノックインが可能である。CRISPR/Cas9 ゲノム編集法は、これまでゲノム編集が難しかったさまざまな生物種で、さらには培養細胞系でも利用できる有用な手法として急速に普及している。

CRISPR/Cas9 ゲノム編集法では、非特異的なゲノム切断 (オフターゲット効果) により望まない変異がゲノムに生じる場合がある。オフターゲット効果を防ぎ、標的部特異的なゲノム編集を行うためには、特異性の高いガイド RNA の配列を選択することが重要なポイントになる。

このツールの特徴

CRISPRdirect は、CRISPR/Cas9 ゲノム編集法のガイド RNA 設計を支援するツールである。前述の GGGenome が対応している 200 種以上の生物種のゲノムに対してオフターゲット候補部位を探索することができ、これらの生物種で利用可能な特異性の高いガイド RNA の設計と選択が可能である。

5. 主な発表論文等

[雑誌論文] (計 4 件)

[Naito Y.](#), [Hino K.](#), [Bono H.](#), [Ui-Tei K.](#)

CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120-1123 (2015) 査読あり

DOI: 10.1093/bioinformatics/btu743

[内藤雄樹](#). GGRNA: キーワードや配列から遺伝子を Google のように検索. 実験医学増刊 (今日から使える! データベース・ウェブツール), 78-79 (2014) 査読なし

[内藤雄樹](#). GGGenome: 高速な塩基配列の検索ツール. 実験医学増刊 (今日から使える! データベース・ウェブツール), 102-103 (2014) 査読なし

[日野公洋](#), [内藤雄樹](#), [坊農秀雅](#), [程久美子](#). CRISPRdirect: CRISPR/Cas9 ゲノム編集法のガイド RNA 設計ツール.

実験医学増刊(今日から使える!データベース・ウェブツール),114-115 (2014)
査読なし

〔学会発表〕(計8件)

内藤雄樹, 坊農秀雅. GGRNA/GGGenome:塩基配列データベースをすばやく検索するウェブサーバ. BMB2015(第38回日本分子生物学会年会, 第88回日本生化学会大会 合同大会), 2015年12月1日発表, 神戸ポートアイランド(神戸市)

吉田徳幸, 内藤雄樹, 佐々木澄美, 内田恵理子, 小比賀聡, 佐藤陽治, 内藤幹彦, 井上貴雄. Gapper 型アンチセンスによる相補結合依存的オフターゲット効果の安全性評価手法構築に向けた基盤研究. 日本核酸医薬学会第1回年会, 2015年11月30日発表, 京都テルサ(京都市)

内藤雄樹, 坊農秀雅. ゲノム編集のオフターゲット効果を防ぐための塩基配列検索ツール. NGS現場の会第4回研究会, 2015年7月2日~7月3日発表, つくば国際会議場(茨城県つくば市)

内藤雄樹, 坊農秀雅. CRISPR/Cas9ゲノム編集法のためのウェブツール. 第13回がんとハイポキシア研究会, 2015年6月5日発表, 国立遺伝学研究所(静岡県三島市)

Naito Y., Hino K., Ui-Tei K., Bono H. CRISPRdirect: web-based tool for designing CRISPR/Cas guide RNA with reduced off-target sites.

CRISPR 2015 OXFORD (2015/3/23)
St Hilda's College, Univ. Oxford, UK

内藤雄樹, 坊農秀雅. 統合遺伝子検索GGRNAと高速塩基配列検索GGGenome:塩基配列データベースをすばやく検索するウェブサーバ. 第37回日本分子生物学会年会, 2014年11月27日発表, パシフィコ横浜(横浜市)

内藤雄樹, 坊農秀雅. ゲノム編集のオフターゲット効果を防ぐための塩基配列検索技術. 第4回ゲノム編集研究会, 2014年10月6日発表, 広島国際会議場(広島市)

Naito Y., Bono H. GGRNA and GGGenome: ultrafast search engines for nucleotide sequence database. Genome Informatics 2014 (2014/9/22)
Churchill College, Univ. Cambridge, UK

〔図書〕(計1件)

内藤雄樹(編集), 羊土社, 今日から使える! データベース・ウェブツール達人になるための実践ガイド100, 実験医学増刊(2014)

〔その他〕

ホームページ等

GGRNA (<http://GGRNA.dbcls.jp/>)
遺伝子を Google のように検索できるウェブサーバ
GGGenome (<http://GGGenome.dbcls.jp/>)
比較的短い塩基配列を高速に検索することができるウェブサーバ
CRISPRdirect (<http://crispr.dbcls.jp/>)
CRISPR/Cas9 ゲノム編集法のためのガイド RNA 設計ウェブサーバ

報道発表

内藤雄樹, 日野公洋, 坊農秀雅, 程久美子. 「ゲノム編集のためのガイド RNA 設計ソフトウェア CRISPRdirect を公開」大学共同利用機関法人情報・システム研究機構ライフサイエンス統合データベースセンター/東京大学大学院理学系研究科. 2014年11月21日発表.

6. 研究組織

(1) 研究代表者

内藤 雄樹 (NAITO, Yuki)
大学共同利用機関法人情報・システム研究機構・ライフサイエンス統合データベースセンター・特任助教
研究者番号: 60451829