



2020 年 3 月 25 日

大学共同利用機関法人 情報・システム研究機構
データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター

高速塩基配列検索ソフトウェア GGGenome パッケージ版の商用化について

大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター (DBCLS) の内藤雄樹特任助教らが開発した、高速塩基配列検索ソフトウェア「GGGenome[®] (ゲゲゲノム)」がこのたび商用化につながりました。

GGGenome は、ゲノムや転写産物などの塩基配列を簡便かつ高速に検索することのできるソフトウェアです。ウェブブラウザ上で利用できるツールとして DBCLS より公開されており、商用・非商用を問わず無償で自由に利用することができます (GGGenome ウェブ版, <https://GGGenome.dbcls.jp/>)。しかし GGGenome ウェブ版は、情報管理上企業等の利用者が秘匿配列を検索することが難しいという問題がありました。そこで、GGGenome ウェブ版と同様の検索を企業等の保有するコンピュータ上で実行可能な「GGGenome パッケージ版」が、株式会社レトリバより発売されることになりました。

株式会社レトリバはこれまで GGGenome の内部の文字列検索エンジンを開発しており、この文字列検索エンジンと、DBCLS がオープンソースとして公開する GGGenome のフロントエンドプログラムを利用し、コンテナ型仮想化技術である Docker を用いて GGGenome パッケージ版を開発することにより、このたびの商用化が実現されたものです。

なお、株式会社レトリバによる GGGenome パッケージ版の発売後も、DBCLS の公開する GGGenome ウェブ版は商用・非商用を問わず無償で自由に利用できます。

GGGenome は、JST ライフサイエンスデータベース統合推進事業「統合データベースにおける基盤技術開発とデータベース運用に係る共同研究」による研究開発の成果です。

また、GGGenome パッケージ版の商用化については、AMED 医薬品等規制調和・評価研究事業「ゲノム編集を利用した遺伝子治療用製品の安全性評価に関する研究 (研究開発代表者 内田恵理子, 研究開発分担者 内藤雄樹, 他)」の支援をいただきました。

1. 背景

GGGenome^{※1}は、ゲノムや転写産物などの塩基配列を見落としなく高速に検索することのできるツールとして、DBCLSの内藤雄樹特任助教らが開発、公開している塩基配列検索ソフトウェアです。内部の文字列検索エンジンは株式会社レトリバが開発しており、ゲノムや転写産物などの塩基配列データから生成した接尾辞配列（suffix array）およびFM-indexとよばれる検索インデックスを、高速な補助記憶装置であるSSD上に配置することにより、高速で見落としのない検索を実現しています。GGGenomeはウェブブラウザ上で利用できるツールとしてDBCLSより公開されており、商用・非商用を問わず無償で自由に利用することができます（GGGenomeウェブ版）。また、GGGenomeウェブ版を構成するプログラムのうち上記の内部エンジン以外のフロントエンド部分は、オープンソースとしてDBCLSより公開されています。

GGGenomeを活用した塩基配列検索のひとつの応用として、核酸医薬品やゲノム編集のオフターゲット候補サイトの検索があげられます。核酸医薬品やゲノム編集は、RNAあるいはDNAの特定の配列を直接標的とすることができるため、これまで有効な治療法のなかった疾患に対する新たな治療法として大きな注目を集めています。一方、核酸医薬品やゲノム編集が標的とする塩基配列は長さが約20塩基と短いため、本来の標的と類似した配列をもつ無関係な部位に結合して意図しない影響を与え（オフターゲット効果）、副作用につながるおそれがあります。オフターゲット効果をインシリコに予測する方法として、ゲノムや転写産物のデータベースを塩基配列検索プログラムを用いて検索する方法が提案されていますが、一般的な塩基配列検索プログラムでは検索に見落としが生じる場合があります。課題のひとつとなっていました。GGGenomeはこのような検索の見落としがないため、核酸医薬品やゲノム編集のオフターゲット候補サイトの検索に最適と考えられます。GGGenomeはすでに製薬企業や公的研究機関において活用されており、初のsiRNA医薬品として2019年に国内で承認された核酸医薬品「オンパットロ」（一般名：パチシラン）の承認審査においても、当該品目の安全性評価にGGGenomeによるオフターゲット候補サイトの検索が貢献しました^{※2}。

一方、これまでGGGenomeウェブ版を利用する場合、配列の検索はDBCLSの保有するサーバ上で実行されるため、製薬企業等が秘匿情報を含む配列を検索することが情報管理の点で難しいという問題がありました。そこで、GGGenomeウェブ版と同様の検索を製薬企業等の保有するコンピュータ上で実行可能な「GGGenomeパッケージ版」の開発が求められていました。

2. GGGenome パッケージ版の概要

株式会社レトリバの開発したGGGenomeパッケージ版は、これまで同社がGGGenomeの内部エンジンとして開発してきたプログラムと、DBCLSの内藤雄樹特任助教らが開発しオープンソースとして公開するフロントエンドプログラムとを、コンテナ型仮想化技術であるDockerを用いてUSBメモリ型のポータブルSSD内にパッケージ化したものです。Mac、Windows、Linux上で動作し、OS等の環境によらず従来のウェブ版と同様の結果が企業等の保有するコンピュータ上

で得られます。オフライン環境でも利用できるため、検索内容の漏洩を防止することが可能です。

また GGGenome パッケージ版（創薬パック）には、創薬研究に必要なヒト、カンクイザル、マーモセット、マウスの転写産物（mRNA および pre-mRNA）のデータベースとして理化学研究所の整備する創薬に資する霊長類のオミックスデータベース D3G^{※3}が含まれており、すぐに検索を実施することができます。

GGGenome パッケージ版に関する詳細は、株式会社レトリバのウェブページをご覧ください。

<https://gggenome.retrieva.jp/>

3. 関連リリース

- ・2020年3月25日 株式会社レトリバ

「高速塩基配列検索ソフトウェア GGGenome（ゲゲゲノム）パッケージ版をリリース」

https://retrieva.jp/info/press/date_202003251100/

- ・2020年3月25日 国立研究開発法人理化学研究所

「創薬に資する霊長類のオミックスデータベース D3G の公開」

https://www.riken.jp/pr/news/2020/20200325_2/

※1 GGGenome は大学共同利用機関法人情報・システム研究機構の登録商標です。

※2 医薬品医療機器総合機構, 「オンパットロ点滴静注 2mg/mL」 審査報告書 (2019)

※3 D3G: Database for Drug Development based on Genome & RNA sequences. <https://d3g.riken.jp/>

問い合わせ先

研究開発について

大学共同利用機関法人情報・システム研究機構
データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター

特任助教 内藤 雄樹

電話：04-7135-5508 E-mail：y-naito@dbcls.rois.ac.jp

報道について

大学共同利用機関法人情報・システム研究機構
データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター

広報担当

電話：04-7135-5508 FAX：04-7135-5534

E-mail：public_relations@dbcls.rois.ac.jp