



2024年4月1日

国立大学法人東京大学医科学研究所

東大医科研ヒトゲノム解析センターがゲノム研究に最適な データサイエンスコンピューティングシステム

Shirokane7 を稼働

——最先端の解析資源と大量データの長期保存環境を融合した次世代生命科学
データ解析基盤を実現——

■発表のポイント

- 国立大学法人東京大学医科学研究所（所長：中西 真／以下、東大医科研）附属ヒトゲノム解析センター（センター長：井元 清哉／以下、HGC）は、最新型のデータサイエンスコンピューティングシステム Shirokane7（以下、Shirokane7）を2024年4月に稼働しました。
- 複数のシステムを統合した生命科学データサイエンス用スーパーコンピュータシステム SHIROKANE（以下、SHIROKANE）（注1）は、Shirokane7 の稼働により最新の AMD CPU や Arm CPU が加わり CPU 総コア数は稼働前に比べ約 1.5 倍の 30,516 コアに達し、最新の GPGPU NVIDIA H100 が使用可能になるなど計算資源を強化しています。
- 国の進める全ゲノム解析等実行計画 2022（注2）にて示されている通り、ゲノム情報を用いた個別化医療を発展させる研究開発の社会に果たす役割は一層重要なものとなっています。東大医科研 HGC は今後も利用者目線にたったデータ解析基盤の提供を通して生命医科学研究の発展に貢献していきます。

■概要

HGC は SHIROKANE のゲノム情報を用いた解析基盤強化のため、新世代システム Shirokane7 を4月1日に稼働し、同日から利用者へ提供を開始しました。Shirokane7 は株式会社日立製作所が構築し、運用期間中における稼働を支援します。Shirokane7 では、Shirokane5 に比べ1ノード当たり 5.3 倍のコア数を備える AMD EPYC シリーズの最新世代（第4世代）CPU を搭載した計算サーバ群を導入するほか、Arm 搭載計算サーバを大幅に拡充しました。さらに、最新の GPGPU NVIDIA H100 を導入し創薬分野をはじめとする AI 技術を用いた新たな研究開発を後押しします。また、テープストレージ機器を刷新し、最大 210 PB のデータを長期間安定して保存する大容量データアーカイブ環境を提供します。



■背景・課題

ゲノム情報を用いた研究は 2003 年のヒトゲノム計画の完了以来、かつてない脚光を浴びています。その大きな要因は、2019 年の新型コロナウイルス感染症パンデミックに対応するなかでゲノム解析を用いた研究の重要性が再認識されたこと、時を同じくして国家戦略として個別化ゲノム医療を大きく発展させるための全ゲノム解析等実行計画が立案されたことがあげられます。東大医科研は前述の全ゲノム解析等実行計画に基づく研究体制の中で、2021 年度に収集されたがん領域約 10,000 症例の全ゲノムデータ解析を約 300 日で完了させた実績を有しています。この全ゲノムデータは、がんゲノムのシーケンス深度を従来の 3 倍に増加させており、最終的な解析結果の正確さと網羅性を向上させ、癌細胞におけるより多くのゲノム異常の検出を可能としました。SHIROKANE は全ゲノム解析等実行計画を推進する専用基盤を支援する形でこの解析実現に貢献しています。

SHIROKANE への期待値・ニーズは、その利用者数や計算ジョブ数の増加として現れています。SHIROKANE は学術機関だけでなく民間機関からの利用も受け付けており、利用者はこの 1 年で約 1.2 倍となり 3,500 を超えています。利用者の増加に伴い同時ジョブ実行数は平均 14,000 を超え計算資源の最大化が急務となっていました。また、Parabricks を始めとする GPU による生命科学分野の高速データ処理が期待されるバイオツールも大きく発展し多様な解析が可能となっており、それらを実行する高性能な GPU サーバのニーズが急速に高まっています。

■今回の取り組み

Shirokane7 では AMD の EPYC シリーズの最新世代である第 4 世代モデルを搭載した計算サーバを導入しました。検証環境にて伝統的な解析ソフトウェア BLAST、BWA、SAMTools sort を対象に実行完了までの時間を測定した結果、Shirokane5 から 30~50% の時間短縮となり、より高速な解析環境を実現しました。

また、2022 年 (Shirokane6) から部分的に導入していた Arm 搭載計算サーバを大幅に拡充しました。Arm アーキテクチャ CPU は他の CPU と比べ、消費電力が低く解析ソフトウェアにおいても同等の実行性能が確認できていますが、Arm CPU のために再コンパイルすることが必要となります。HGC では、利用頻度の多いバイオツールから再コンパイルを実施しており、SHIROKANE で利用可能なツールは 2024 年 4 月時点で 90 ツールを超えています。Shirokane7 稼働後は、無料コースと試用コースを除く全てのコースに 11,488 スロットの計算ノード Arm を提供し、追加手続きなく新しい計算ノード Arm が使用可能です。

HGC では、生成 AI 技術の発展と創薬分野への技術適用の期待を背景に、NVIDIA BioNeMo™ を始めとした NVIDIA NGC™ で提供されているフレームワークなどを SHIROKANE で実行できる GPU 環境を整備しています。NVIDIA BioNeMo™ は創薬のための生成 AI プ



ラットフォームであり、AI 基盤の新たな創薬パイプラインとして期待が高まっています。利用者は必要に応じて NVIDIA BioNeMo™ Framework を始めとする生成 AI 技術を用いた解析が実行可能です。

■今後の取り組み

HGC は、SHIROKANE を最先端ゲノム研究の基盤とし、全ゲノムシーケンスデータをはじめとするマルチオミクスデータ解析や最先端の大規模言語モデルなどの AI を用いた大規模解析が可能な環境を提供していきます。利用者目線に立った質の高いサービスを提供することにより、公共・民間問わず国内の生命科学研究を大きく加速させ、医学の発展と社会に貢献していきます。

■商標に関する表示

記載の会社・組織名、製品名は、それぞれの会社・組織の商標もしくは登録商標です。

■用語解説

(注1) 2024 年 4 月からの SHIROKANE は、Shirokane5 (2019 年 4 月から運用開始)、Shirokane6 (2022 年 4 月から運用開始) そして Shirokane7 で構成される。

(注2) 国の進める全ゲノム解析等実行計画 2022 については、厚生労働省ホームページを参照下さい。 <https://www.mhlw.go.jp/content/10808000/000996722.pdf>

■本件に関する問合せ先

国立大学法人東京大学医科学研究所 附属ヒトゲノム解析センター

教授・センター長 井元 清哉 (いもと せいや)

https://www.ims.u-tokyo.ac.jp/imsut/jp/lab/hgclink/page_00072.html

■報道に関する問合せ先

国立大学法人東京大学医科学研究所 プロジェクトコーディネーター室 (広報)

<https://www.ims.u-tokyo.ac.jp/>