

2020年3月10日
国立大学法人東京大学
株式会社日立製作所

**東大医科研と日立がヒトゲノム解析用スーパーコンピュータ Shirokane5 を用いた
ゲノムデータ解析の高速化に向けた検証において、解析時間を従来比約 80%削減**
がんゲノム医療におけるゲノム変異箇所の検索・特定時間を短縮し、
Society5.0 時代の個別化医療の実現に貢献

■本発表のポイント

- ・ 国立大学法人東京大学 医科学研究所 ヒトゲノム解析センターは、株式会社日立製作所の協力のもと、最新型のヒトゲノム解析用スーパーコンピュータシステム Shirokane5 を用いて、がんゲノム医療における全ゲノムデータ解析の高速化に向けた検証を実施し、解析時間を従来比約 80%削減することに成功しました。
- ・ 今後、高度なデータ解析を高速に処理できる高セキュリティな Shirokane5 の環境において、本格的に全ゲノムデータ解析を開始します。
- ・ 膨大なゲノムデータからがんの原因となったゲノムの変異箇所を検索・特定するまでを高速化し、一人ひとりの体質や病態にあった治療方針を迅速に提供するなど、Society5.0 時代のゲノム情報を活用した個別化医療の実現に貢献します。

■概要

国立大学法人東京大学 医科学研究所(所長: 山梨 裕司/以下、東大医科研)ヒトゲノム解析センターは、株式会社日立製作所(執行役社長兼 CEO: 東原 敏昭/以下、日立)の協力のもと、最新型のヒトゲノム解析用スーパーコンピュータシステム Shirokane5 を用いて、がんゲノム(注 1)医療における全ゲノムデータ解析の高速化・解析時間の短縮化に向けた検証を実施し、従来 10 時間以上を要していた解析時間を最短 1 時間 45 分で解析するなど、従来比約 80%削減することに成功しました。

これにより、東大医科研 ヒトゲノム解析センターは、個人のヒトゲノムの特徴に応じたがんや生活習慣病などの予防・診断・治療法の研究を加速し、Society5.0(注 2)時代のゲノム情報を活用した個別化医療の実現を支援します。

なお、日立は、2019 年 4 月に、高セキュリティな Shirokane5 を設計・構築し、その後の運用を支援してきました。また、今回の検証にあたり、Shirokane5 上にゲノムデータ解析プログラムを実装し、全ゲノムデータを高速で処理できる高度な解析環境を構築するなど、ゲノム解析の高速化に向けた取り組みを支援しています。

■背景・課題

一人ひとりの体質や病態にあった適切な治療を提供するために、ゲノム情報を活用した医療(ゲノム医療)の発展が期待されており、日本においても、がんを対象としたゲノム医療が、保険収載のもとで 2019 年 6 月に開始されました。がんゲノム医療では、がん細胞のゲノムを調べることによって、がんの原因となったゲノムの変異を突きとめ、その情報からがんの弱点を見いだすことで効果が期待される薬剤を選択し、一人ひとりのがんの特徴にあったより効果の高い治療を提供することを目指しています。

現在、日本のがんゲノム医療では、これまでの研究からがんとの関係性が明らかになっている数百規模の遺伝子のゲノム配列を調べることができる「遺伝子パネル検査」が使われているものの、ヒトには約 2 万 1 千個の遺伝子があるため、全ての遺伝子は調べられていないことになります。また、仮に全遺伝子を調べたとしてもそれらの占める領域は、全ゲノム領域のたった 2%程度に過ぎず、ゲノム上には「遺伝子パネル検査」では調べていない領域が膨大に存在します。

近年、最先端のゲノム研究によって遺伝子以外のゲノム領域に生じた変異もがんを生じさせる原因となることも明らかになってきており、将来のがんゲノム医療には、全ゲノム情報を調べることが必須となると考えられます。しかしながら、全ゲノムの膨大な情報を解析することは、従来 of 大型計算機を使っても膨大な時間がかかる作業でした。そのため、全ゲノム情報に基づくゲノム医療を多くの患者に提供するための情報解析基盤の構築が課題となっています。

■ 今回の検証内容と成果

今回、ゲノムデータ解析の高速化に向け、東大医科研 ヒトゲノム解析センターの有する最新型のヒトゲノム解析用スーパーコンピュータシステム Shirokane5 を活用し、日立の協力のもと、全ゲノムデータの解析時間の短縮に向けた検証を行いました。本検証は、Shirokane5 やゲノムデータ解析プログラムを用いた高度な全ゲノム解析を高速に実行可能な環境でゲノム解析を行い、東大医科研でこれまでに開発してきた、がんゲノム医療のワークフローに組み込むことにより、従来 10 時間以上を要していた解析時間を約 80%削減し、最短 1 時間 45 分での解析に成功しました。

これにより、患者からゲノム解析の同意を受けてから、ゲノム医療の結果を担当医に返却するまでに要する最短期間を、従来の 3 日と 7 時間 30 分から 2 日と 16 時間 へ大幅に短縮可能としました。

ある種の白血病(血液のがん)は、週単位、時には日々病態が変化し急激に悪化することもあり、一日でも早くがんの原因となったゲノム変異を特定し、効果的と予測される治療方針を立てることが医療の効果を高める鍵となります。また、他のがんにおいても迅速にゲノム解析結果を返却できることは、より早期の治療開始を可能にすることから、患者不安を軽減し、治療の効果を高めることが期待されます。

現在、科学研究において、全ゲノム情報を用いた成果が次々に出ているものの、その成果を医療に還元するにあたり、膨大な情報解析の負担が実現を阻む障壁の一つでした。本検証により、多くの全ゲノム情報解析を短時間で実施可能とする成果を得られたことは、全ゲノム情報を最大活用した将来のがんゲノム医療の実現に大きく貢献するものです。

なお、今回の検証は、Shirokane5 に搭載している GPU(注 3)全 80 基のうち 16 基で実施したため、全 GPU を稼働させた場合、1 年間で 7 万検体の全ゲノムデータ解析が可能となります。これは、がん細胞と正常細胞の 2 つの解析を行うがんゲノム医療において、がん患者数にすると約 3 万 5 千人分の全ゲノムデータ解析を可能とすることに相当します。

日立は、長年にわたり、ゲノム解析の高速化に向け東大医科研の取り組みを支援してきました。2019 年 4 月に Shirokane5 を設計・構築し、Shirokane5 に搭載する GPU 搭載サーバ上に、NVIDIA 社のゲノムデータ解析ソフトウェア Parabricks を短期間で実装し、パフォーマンスの検証や最適化などを行い、高度な全ゲノムデータ解析を高速処理できる環境を実現しています(注 4)。

また、本環境はヒトゲノム所有者の個人情報であるパーソナルゲノム情報を安全に扱うため、ヒトゲノム情報を扱うシステム領域を外部のインターネットと分離し、迅速なシステム変更能耐えられる構成にすることなどにより、柔軟かつセキュリティ性の高いシステム環境を実現しています。

■今後の取り組み

東大医科研 ヒトゲノム解析センターは、今後、本解析システムを活用し、最先端のがん研究にも応用し、猛烈なスピードで進展するがん研究においても世界をリードするゲノム解析研究を推進するとともに、解析基盤としての支援も持続させ、さらなるゲノム医療の発展に貢献します。

日立は、これまで培ってきたゲノム解析基盤やオープンソースソフトウェアの構築経験を生かしてヒトゲノム研究の推進を支援し、国民の健康・医療の向上やライフサイエンス分野の進展、さらには Society5.0 時代のゲノム情報を活用した個別化医療の実現に貢献します。

■がんゲノム医療のワークフローについて

がんゲノム医療のワークフローは、大きく 4 つで構成され、(1)ゲノムの読み取り(シーケンス解析)、(2)がん細胞に生じたゲノム異常を特定、(3)特定された大量のゲノム異常の臨床的意義を AI の支援により解釈し、(4)効果の期待できる薬剤を予測します。

今回の検証では、ボトルネックであった(2)を構成するシーケンスデータのアライメント(注 5)と変異の特定と呼ばれるデータ解析作業にかかる時間を大幅に短縮することに成功しました。



■スーパーコンピュータシステム Shirokane5 の特長と構成

1. 高度なビッグデータ解析を高速に処理

Shirokane5 では、1.78PFLOPS(ペタフロップス(注 6))の総合理論倍精度演算性能(注 7)を実現しています。具体的には、ヒトゲノム情報の高速な並列処理を行うプログラム用には合計 13,176 CPU コアを有する「分散メモリ型サーバ」、AI による高速なゲノム解析を行うプログラム用に合計 NVIDIA Tesla V100 80 基を有する「GPU 搭載サーバ」、ヒトゲノムアセンブリプログラムといった大容量のメモリを必要とするプログラム用には 1 ノードあたり 3 TB のメモリ容量を有する「大規模メモリサーバ」を有しています。

2. 高セキュリティ環境の安定稼働を実現

ヒトゲノムを扱った研究においては、ヒトゲノム所有者の個人情報と厳重に管理するため、セキュリティを確保した上で、研究者やシステム運用者が所有者個人を特定することがないよう、必要に応じて匿名化の手続きを行っています。あわせて、東大医科研 ヒトゲノム解析センターでは、ヒトゲノム情報を扱うシステム領域を外部のインターネットと論理的に隔離し、ログイン認証に生体認証を採用する高セキュリティ領域を確保しています。

3. 電力消費の効率化

Shirokane5 では環境に配慮したシステム運用を行っています。測定した電力、冷却装置の各ファンの電力、計算機の動作状況、温度・湿度などの多様な情報を集め、蓄積・表示、そして冷却装置の制御に自動フィードバックするシステムを作成しました。これにより、冷却装置の過剰な動作を抑止し、消費電力の削減を実現しています。現在は、蓄積した稼働情報からニューラルネットワーク(注 8)を用いて学習させ、より最適化を図り、さらなる効率化に向けて取り組んでいます。

注 1: ゲノムとは、遺伝子をはじめとした遺伝情報の全体を意味する。また、がんゲノム医療は、遺伝子情報に基づくがんの個別化治療の 1 つ。

注 2: 日本政府が掲げる新たな社会像であり、その実現に向けた取り組みのこと。AI や IoT、ロボットなどの革新的な科学技術を用いて、社会の様々なデータを活用することで、経済の発展と社会課題の解決を両立し、人間中心の豊かな社会をめざす。狩猟社会、農耕社会、工業社会、情報社会に続く 5 番目の新たな社会として位置づけられている。

注 3: GPU(Graphics Processing Units): 高度な画像処理を行うためのプロセッサ。1999 年に NVIDIA 社が世界ではじめて開発。高度な並列演算性能を備えており、AI(ディープラーニング)や科学技術計算などの HPC に活用される。

注 4: コンテナ型仮想化技術を用いることで、変化の早いゲノム解析市場において短期間での構築を可能にしたほか、多数の利用者が共用する Shirokane5 において他ソフトウェアに依存しない独立した環境での構築を可能にしソフトウェアの信頼性を向上。なお、コンテナ型仮想化は、従来のハードウェア仮想とは異なり、一つの OS を複数の仮想環境が共有して利用するため、ソフトウェアでハードウェアを代行する方式よりも起動や動作が速くなるもの。

注 5: シークエンサーから出力される断片化されたゲノム配列を既知の参照ゲノムと照らし合わせ整列しなおす作業。

注 6: PFLOPS(ペタフロップス): 浮動小数点演算を 1 秒間に 1000 兆回実行する能力。

注 7: 総合理論倍精度演算性能: 同時に動作可能な全ての演算器が動作したときの理論上の性能。

注 8: 人間の神経回路網の動作を模した処理モデル。

■ 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。

■ 本件に関するお問い合わせ先

国立大学法人東京大学 医科学研究所 ヘルスインテリジェンスセンター

教授・センター長 井元 清哉 (いもと せいや)

〒108-8639 東京都港区白金台 4-6-1

URL: <http://www.ims.u-tokyo.ac.jp/imsut/jp/>

株式会社日立製作所 公共システム営業統括本部

カスタマ・リレーションズセンタ [担当: 森下]

〒140-8512 東京都品川区南大井六丁目 23 番 1 号 日立大森ビル

URL: <https://www.hitachi.co.jp/public-it-inq/>

以上