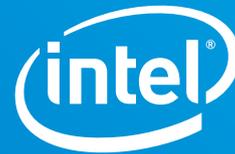


ソリューション概要

ライフサイエンス

Falcon Computing によるゲノム解析
ソフトウェアとインテル® Arria® 10 FPGA



Falcon Computing ソフトウェアと インテル® Arria® 10 FPGA による ゲノム解析の高速化

ゲノム解析における高い精度と再現性を維持しながら、
結果算出の時間を短縮する Falcon Computing ソリューション

「研究および

精密医療の分野では、
結果の精度と再現性を
確保するために、
業界標準のパイプライン上での
コスト効率の高い、
ハイパフォーマンスの解析が
求められます。私たちは
GATK パイプラインの高速化に
積極的に取り組み、
最大 15 倍のスピードアップを
実現しています」¹

— Falcon Computing
製品およびマーケティング担当副社長
Sunder Parameswaran 氏

概要

ゲノム解析は、生体からの DNA 配列データの生成と、ゲノムの遺伝子配列へのマッピングから構成されています。この解析作業は、医療における技術革新、パーソナライズ医療、さらには希少遺伝性疾患の診断および治療にとって非常に重要です。精密医療、ゲノミクス、エピジェネティクスでは、ゲノム解析を使用することで、医療サービス提供者向けの研究の実施、診断の向上、薬の開発、治療の質の向上を行うとともに、農作物の分野でも生産の最適化を行っています。

解析のコストはここ数年で劇的に低下しましたが、データは複数のソースから生成されるために処理や統合という課題が生じています。Falcon Accelerated Genomics Pipeline (FAGP) は、インテル® Arria® 10 FPGA によって高速化され、インテル® プロセッサによって強化されたエンドツーエンドのソフトウェア・ソリューションで、ゲノム解析の高速化と簡素化を実現しています。インテルとの連携により、Falcon Computing は最大 15 倍の高速化を実現するプラグ・アンド・プレイ機能をオンプレミスとクラウドの両方で提供しています。これにより、ゲノム解析をより短時間で処理できるとともにコストが低減され、インサイトを得るまでの時間が短縮されます。¹

課題

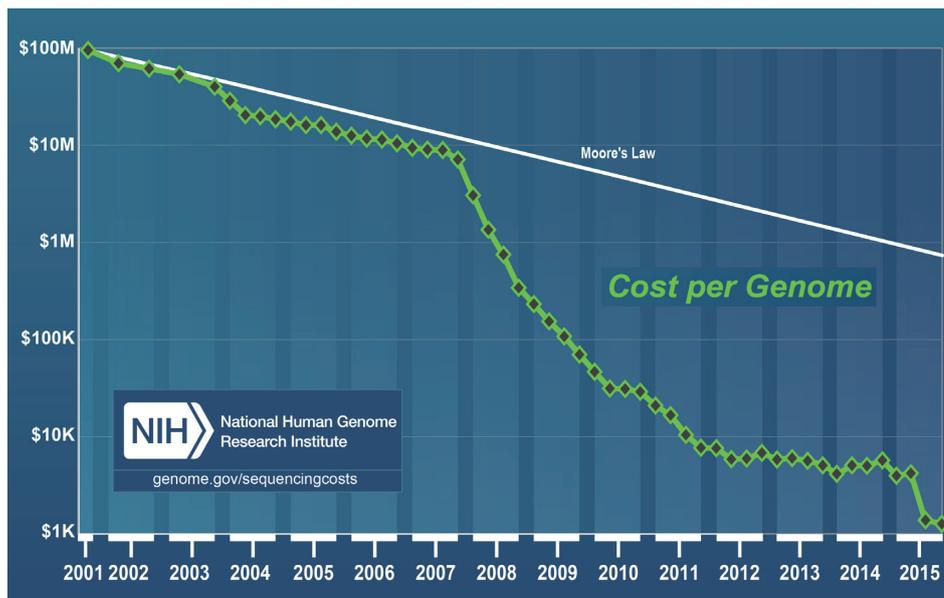
1 人当たりのヒトゲノムのデータ量は数百 GB にも達します。さまざまな研究目的や診断目的でこうしたデータが数百万人分もの規模で生成され、大容量のデータへと変換されています。ヒトゲノム・プロジェクトが開始されて以来、ライフサイエンス業界はデータの飛躍的な増加の時代を迎えています。ライフサイエンスの領域では、次世代シーケンシング (NGS) アプライアンスによるデータ収集の大幅なコスト削減により、全ゲノムの解析が現在重要なアプリケーションとなっています。データの増加に加えて、大学、ゲノム研究センター、製薬会社、医療組織で使用されるゲノム・アプリケーションの範囲も拡大しています。

ヒトゲノム・プロジェクトの開始以来、ここ 20 年でゲノム解析のコストは、1 ゲノム当たり 1 億ドル以上から 1,000 ドル未満まで減少しました。こうしたコスト削減によって、世界中で解析されたゲノムの数は飛躍的に増大しました。実際、ゲノムデータの量は 7 か月ごとに 2 倍になっています。その結果、データ処理を効率化し、コスト面でも効率の高い方法で行うことが重要となってきました。しかし、一般に CPU の速度が 2 倍になるには 2 年かかると言われ (ムーアの法則)、CPU の演算能力の向上だけでゲノムデータの増加に対応することは難しく、ハードウェア・アクセラレーションが必要になります。FPGA などのアクセラレーターは、このゲノムデータの急増による演算ニーズに対応するための切り札となりつつあります。

現在、業界では、コスト効率の高いゲノム解析を業界標準のパイプライン上で提供することが求められています。また、これを多様なクラウドおよびハードウェアを対象として実現するとともに、結果について臨床レベルの精度と再現性を達成しなければなりません。



ハードウェア・アクセラレーションによってゲノムデータ増加のニーズに対応可能



出典: <http://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/> (英語)

DNAデータは7カ月ごとに2倍になるのに対し、CPUの速度が2倍になるには2年かかる

ソリューション

Falcon Accelerated Genomics Pipeline (FAGP) が提供する、コンフィグレーション済みのエンドツーエンドのソフトウェア・ソリューションは、GATKパイプラインを使用してゲノムデータを高いコスト効率で解析し、高いパフォーマンス、精度、再現性を実現します。このソリューションは最大 15 倍のスピードアップと、GATK上で 100%の精度を実現します。¹ つまり、従来 50 ~ 60時間かかっていた解析を4時間未満で完了できるようになります。FAGPは、信頼性に優れたハイパフォーマンスのインテル® Arria® 10 FPGAおよびインテル® Xeon® プロセッサと併用することで、卓越したアクセラレーションと精度を提供します。さらに、FAGPは、全ゲノム解析を2時間以内に完了できる、より高速なパイプラインも提供します。

ゲノム解析向けのハイパフォーマンス統合ソリューション

GATKパイプラインを使用して、全ゲノム解析を4時間未満、エクソーム解析を30分未満で実施します。¹

- 独自規格ではなく、業界標準のGATKパイプライン上での高速化
- インテル® FPGA上の複数GATKコンポーネントのアクセラレーションを介した優れたパフォーマンス
- あらゆる主要NGSワークフローとの互換性
- リードのローデータをVCFファイルに正確に変換
- インテル® プログラマブル・アクセラレーション・カード (インテル® Arria® 10 GX FPGA搭載版) およびインテル® アクセラレーション・スタック (インテル® Xeon® CPU & FPGA対応) を使用したアクセラレーションによる高いスループットと効率

「TCGBでは、ゲノム解析サービスを全米のクライアントに提供しています。

Falcon Accelerated Genomics Pipelineによって、業界標準のGATKパイプラインの精度を維持しながら、所要時間を数日から数時間へと短縮することができました」

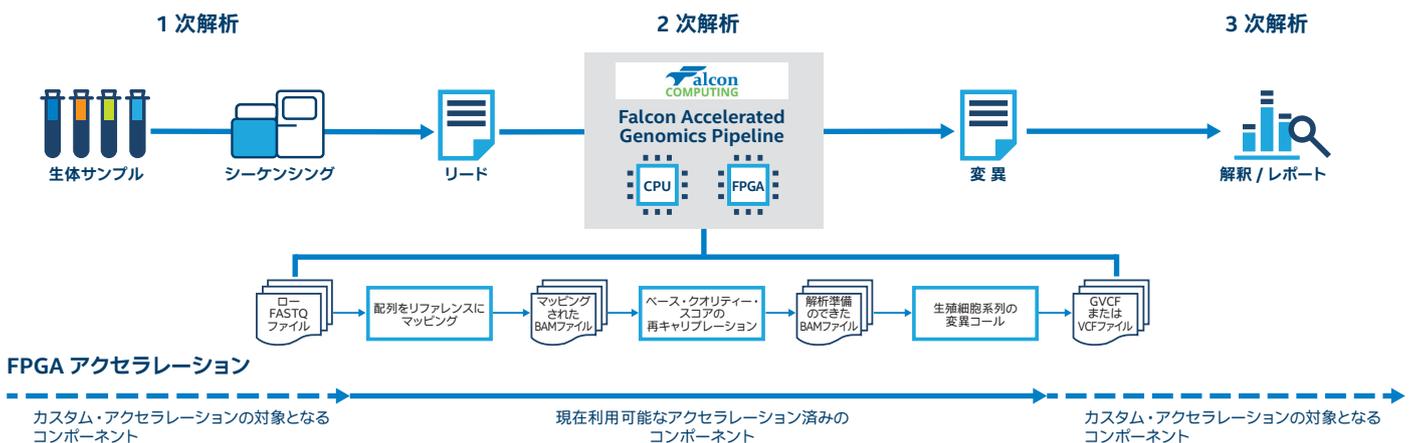
— UCLA, Technology Center for Genomics & Bioinformatics (TCGB)、ディレクター
Xinmin Li 博士

このプラグ・アンド・プレイ・ソリューションは、クラウドでもオンプレミスでも、対象のインテル® プロセッサ搭載サーバー上で容易にインストールし、シームレスに実行することができます。FAGPのコンフィグレーションはユーザー・フレンドリーなコンソール経由で行われるため、データ・サイエンティスト、研究者、大学関係者、製薬会社やシーケンシング・サービス・プロバイダーにおけるワークフローへの統合も容易です。

Falcon Computingにより、ゲノム解析を高いコスト効率で複数のエンドポイントにわたって繰り返し導入し、常に同じ結果を生み出すことができます。Falcon Computingのソフトウェア・アプライアンスはオンプレミスまたはクラウド内のユーザーのサーバー上にインストールされるので、ユーザーはデータの所有権、プライバシー、制御を確実に維持できます。Falcon ComputingのCost Optimization Engineでは、時間の制約とサンプル数の組み合わせを入力として、処理速度および予算のニーズに合うソリューションを提供します。

FAGPにより、パイプラインの任意のステップへのアクセスやスキップ、パラメーターの変更、調整を可能にする柔軟性がもたらされます。独自規格のカスタム・パイプラインを内製している場合に、Falcon Computingは、Merlin Compilerを含むFalcon Accelerationプラットフォームを提供することで、独自パイプラインの高速化を可能にします。さらにFalcon Expert Servicesにより専門知識も提供し、アクセラレーションのニーズに関する顧客のサポートも行っています。

Falcon Accelerated Genomics Pipeline (FAGP) の優位性	
GATKとの完全互換性	<ul style="list-style-type: none"> 複数のGATKバージョン(4.0など)のサポート 生殖細胞系列および体細胞系列パイプラインのベスト・プラクティス 各ステップについてGATKの結果と一致
大規模な業界データの処理能力	<ul style="list-style-type: none"> 6つの全ゲノム解析または24の全エクソーム解析を1日で実行 処理可能なサンプル量の向上 オンプレミス・ソリューションの高速化 オンプレミスでもクラウドでも同じ結果を取得
代替の変異コール・パイプライン	<ul style="list-style-type: none"> WGSについてオンプレミスで3時間未満の所要時間(50倍) 臨床ニーズ向けに高速で正確な変異コール・パイプラインが利用可能 FPGA上での複数のコンポーネントのアクセラレーションにより使用率を最大化
インサイト獲得までの時間の短縮	<ul style="list-style-type: none"> クラウドでもオンプレミスでもベスト・プラクティスのGATKパイプラインを最大15倍高速に実行 インテル® Arria® 10 FPGAベースのアクセラレーション・プラットフォーム上で並列化アルゴリズムにより実行時間を最大35倍改善
既存のコードベースの活用	<ul style="list-style-type: none"> 稼働中のアルゴリズムの再作成が不要 既存のワークフローを最小限のコード変更で使用可能 カスタム・パイプラインのアクセラレーションのための拡張性



標準のGATKパイプライン上で精度と速度を向上

ユースケース

Falcon ComputingのFAGPは、研究施設や製薬会社のように独自のデータを所有している組織でも、シーケンシングをサービスとして提供しているプロバイダーのように顧客から収集したデータをシーケンシングしている組織でも使用できます。

FPGAアクセラレーションのメリットがある主なライフサイエンス・ワークロードを以下に示します。

- ゲノミクスおよびプロテオミクス
- 生物工学
- 分子力学シミュレーション
- パイオインフォマティクス
- 医療用画像処理

例えば、Falcon ComputingはUCLAのCancer InstituteおよびTechnology Center for Genomics & Bioinformatics (TCGB)と連携しています。

UCLAのTCGBは、完全に自動化された高スループットのゲノミクス施設であり、主要な次世代シーケンシング・プラットフォームとマイクロアレイ・プラットフォームをすべて備えています。これらのサービスを高いコスト効率でタイムリーに提供するTCGBは、基礎科学、トランスレーショナル・リサーチ、臨床研究の研究者に役立っています。TCGBでは、FAGPを使用してゲノムデータの2次解析を高速化し、それによってUCLAの研究コミュニティに所要時間の短縮を実現しています。

同時に、UCLAの医科大学院では、FAGPを使用して、免疫療法を待っている非小細胞肺癌患者における突然変異の識別を高速化しています。免疫療法は、身体の防御システムを利用して、がんに苦しむ患者の腫瘍の増大と闘う新しい治療形態です。

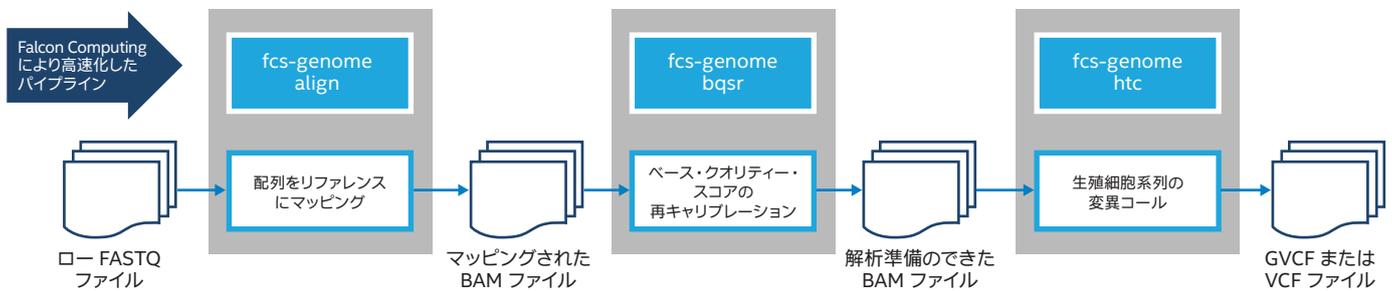
ソリューション・アーキテクチャー

FAGP は、ベスト・プラクティスである GATK パイプラインの高速化バージョンです。このソリューションでは、ローファイルから高速ファイルまでのパイプラインをサポートします。つまり、すべての変換の処理、配列のリファレンスへのマッピング、データのスコアリング、変異の発見といったタスクがサポートされています。

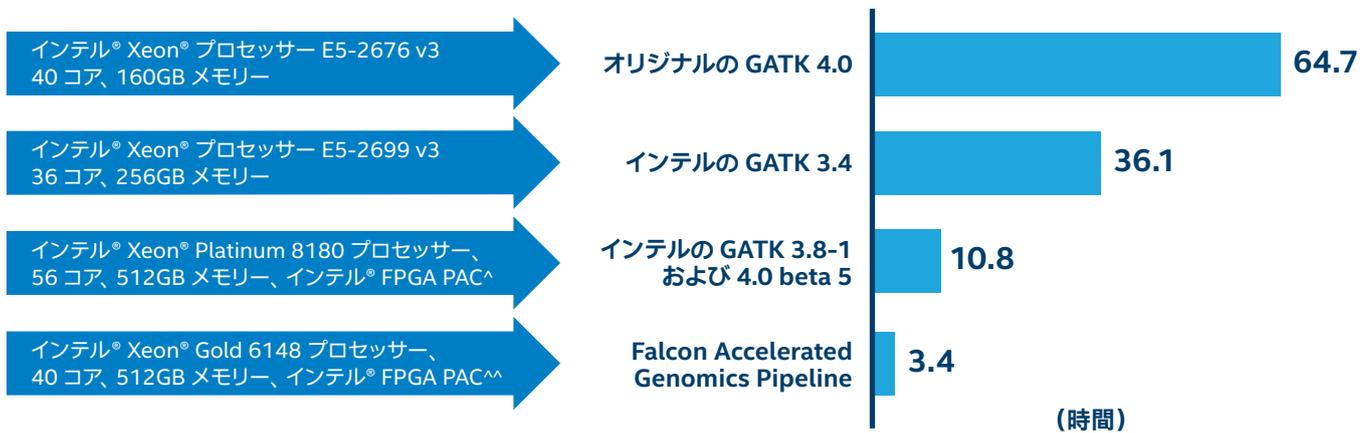
入手後は、コンソール・ダッシュボードからシンプルな3ステップを行うだけで、アプリケーションを使用できます。

1. リファレンスとの位置合わせ
2. ベースの再キャリブレーション
3. 変異コール

パフォーマンスの指標は、ソリューションがローカルで動作するかクラウドで動作するかにかかわらず同じもので、中間ファイルへのアクセスも提供されます。このソリューションは、Falcon Computing の C/C++ から FPGA へのアクセラレーションと併用して、独自規格のカスタム・パイプラインを高速化することもできます。



インテル® アーキテクチャーを搭載した、Falcon Computing が提供する統合ソリューションにより、エンドツーエンドの GATK パイプラインを高速化



[^] 1つのアルゴリズムをFPGAで実行

^{^^} Falcon Computing ソリューションで、複数のアルゴリズムをFPGAで実行

業界標準のパイプラインによる高速シーケンシング：1日で6つの全ゲノム解析または24の全エクソーム解析¹

ソリューションのコンポーネント

- インテル® プログラマブル・アクセラレーション・カード (インテル® Arria® 10 GX FPGA 搭載版)
- 認定サーバーのデュアルソケット・インテル® Xeon® スケーラブル・プロセッサ 6148
- インテル® アクセラレーション・スタック (インテル® Xeon® CPU & FPGA 対応) DCP 1.2
- インテル® FPGA SDK for OpenCL™ AOCL 17.1.1
- Falcon Accelerated Genomics Pipeline (FAGP)
- Genome Analysis Toolkit (GATK)

「GATK4 開発チームの主なミッションは、科学の領域における高度化した課題をできるだけ高い精度で解決するアルゴリズムを作成し、それをゲノミクス・コミュニティ向けのオープンソースのツールとして共有することです。現時点で、非常に多くのチームがこれらのツールの動作を高速化し、時間要件の厳しい GATK4 アプリケーションを実現していますが、これは素晴らしいことです」

— Broad Institute

データ・サイエンス・プラットフォームにおけるアウトリーチ & コミュニケーション担当アソシエイト・ディレクター
Geraldine Van der Auwera 氏

まとめ

ゲノムデータは、ライフサイエンス、製薬、農業などの業界で活用されてきました。インテルと Falcon Computing は、インテル® PAC インテル® Arria® 10 GX FPGA 搭載版とインテル® アクセラレーション・スタック (インテル® Xeon® CPU & FPGA 対応) を使用して、研究および臨床導入の両方を対象として GATK ベースのゲノム処理を高速化しています。

その結果、遺伝性疾患の特定、がんの進行を速める突然変異の特性評価から病気の発生の追跡まで、幅広い成果がもたらされています。

Falcon Computing とインテルによって、より専門的で正確、そして迅速な発見を可能にするための信頼性が高く画期的な道筋が得られます。

詳細情報

Falcon Accelerated Genomics Pipeline (FAGP) の詳細や評価ライセンスの取得方法については、<http://www.falconcomputing.com/> (英語) を参照してください。

インテル® FPGA アクセラレーション・ハブの詳細：
<http://www.intel.co.jp/fpgaacceleration/>

インテル® Arria® 10 FPGA の詳細：
<http://www.intel.co.jp/arria10/>

ライフサイエンスやゲノミクス向けのインテルのソリューション：
<http://www.intel.com/content/www/us/en/healthcare-it/life-sciences.html> (英語)

インテルと Broad Institute のコラボレーション：
<http://www.intel.co.jp/content/www/jp/ja/healthcare-it/solutions/genomics-broad-data.html>

Broad Institute の詳細：
<http://www.broadinstitute.org/> (英語)

Falcon Computing について

Falcon Computing は、ゲノミクス、マシンラーニング、金融、コンピューター・ビジョンの分野で活躍するデータサイエンティストやソフトウェア開発者が、ヘテロジニアス・コンピューティングにおけるアクセラレーションへの筋道を簡素化できるよう支援しています。シンプルな C/C++ から、パブリック・クラウドとプライベート・クラウドの双方に向けて最適化された FPGA 実装まで、ユーザーはパフォーマンス、電力効率、生産性を向上することができます。

<http://www.falconcomputing.com/> (英語)





¹ <http://www.falconcomputing.com/falcon-accelerated-genomics-pipeline/> (英語)

性能の測定結果は2018年1月時点のテストに基づいています。また、現在公開中のすべてのセキュリティー・アップデートが適用されているとは限りません。詳細については、公開されている構成情報を参照してください。絶対的なセキュリティーを提供できる製品はありません。

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサ用に最適化されていることがあります。SYSmark* や MobileMark* などの性能テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行ったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。詳細については、<http://www.intel.com/benchmarks/> (英語) を参照してください。

インテル® テクノロジーの機能と利点はシステム構成によって異なり、対応するハードウェアやソフトウェア、またはサービスの有効化が必要となる場合があります。実際の性能はシステム構成によって異なります。絶対的なセキュリティーを提供できるコンピューター・システムはありません。詳細については、各システムメーカーまたは販売店にお問い合わせいただくか、<http://www.intel.co.jp/> を参照してください。

Intel、インテル、Intelロゴ、Arria、Xeonは、アメリカ合衆国および/またはその他の国におけるIntel Corporationまたはその子会社の商標です。

* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

©2019 Intel Corporation. 無断での引用、転載を禁じます。

OpenCL および OpenCL のロゴは Apple Inc. の商標であり、Khronos の許可を得て使用しています。

338686-001JA