FICHA DE DADOS

DRAGEN™ secondary analysis

Identificação de variantes precisa, abrangente e eficiente para dados de sequenciamento de última geração

illumina®

Somente para pesquisa. Não deve ser usado para procedimentos de diagnóstico.

Introdução

Liberar o potencial do genoma por meio do sequenciamento de última geração (NGS) é fundamental para o avanço da pesquisa biomédica e da medicina de precisão. Para maximizar as descobertas genéticas do NGS, os pesquisadores precisam de ferramentas de análise de dados que possam traduzir com precisão e eficiência dados brutos de sequenciamento em resultados significativos. Além disso, para aproveitar os benefícios do NGS, as organizações requerem soluções fáceis de usar, que acomodem uma ampla variedade de usuários e cuja adoção tenha menores barreiras financeiras e técnicas.

O DRAGEN (Dynamic Read Analysis for GENomics) secondary analysis da Illumina foi desenvolvido para abordar desafios importantes na análise de dados NGS em uma ampla variedade de metodologias, incluindo sequenciamento de genoma completo, sequenciamento de exoma, análise de transcriptoma, ensaios de metiloma e muito mais. O software DRAGEN secondary analysis é um pacote de aplicações que processa dados de NGS e permite a análise terciária para gerar insights. As ferramentas disponíveis constituem uma solução altamente precisa, abrangente e eficiente, permitindo que laboratórios de todos os tamanhos e disciplinas aproveitem mais seus dados genômicos.

Resultados precisos

O DRAGEN secondary analysis gera resultados excepcionalmente precisos. Na competição Precision FDA Truth Challenge V2 (PrecisionFDA V2) de 2020, o DRAGEN secondary analysis v3.7 foi mais preciso em todas as regiões de referência e regiões difíceis de mapear com dados de sequenciamento da Illumina.[1,2] As versões subsequentes continuam a definir novos padrões de precisão, com avanços em áreas como aprendizado de máquina (ML) e tecnologia multigenômica DRAGEN.* A versão mais recente, DRAGEN secondary analysis v4.3, fornece precisão de identificação de variantes pequenas sem precedentes com uma pontuação F1 de 99,89% (uma medida combinada de precisão e recall) em todas as regiões de referência (Figura 1). Isso é possibilitado pelo mapeamento multigenômico DRAGEN de última geração, feito com base em 128 amostras com 256 haplótipos de dados de referência pangenômica construídos internamente, capturando maior diversidade genética.

* O termo "referência multigenômica (gráfica)" foi usado anteriormente para se referir tanto ao método de mapeamento quanto à coleta de amostras de referência. Para descrever melhor as atualizações de versão do DRAGEN secondary analysis, agora são usados termos distintos. "Mapeamento multigenômico" refere-se ao método de mapeamento, e "referência pangemônica" refere-se à coleta de amostras de referência.

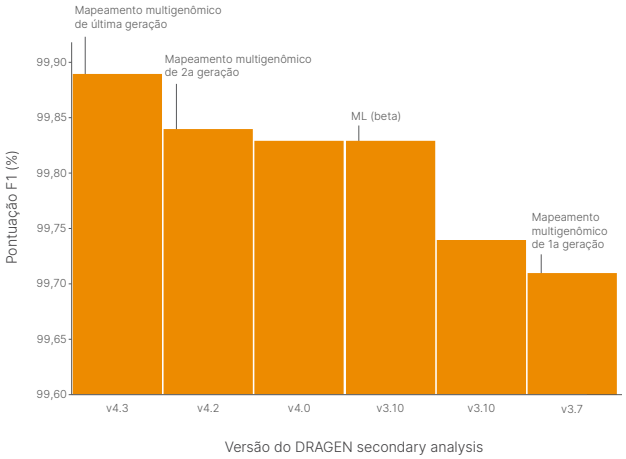

Figura 1: Precisão do DRAGEN secondary analysis: a pontuação F1 do eixo Y (%) é um cálculo de resultados positivos e negativos verdadeiros como uma proporção dos resultados totais.[3,4]

O identificador integrado de mosaicos também contribui para uma maior precisão, podendo ser habilitado para detectar variantes de mosaico com frequências de alelos baixas de até 3%.

Análise abrangente

Com cobertura abrangente do genoma e um amplo conjunto de aplicações compatíveis, o DRAGEN secondary analysis atende às diversas necessidades dos laboratórios que fazem análise de NGS. Os pipelines do DRAGEN são compatíveis com vários tipos de experimentos, incluindo sequenciamento do genoma completo (WGS), sequenciamento do exoma completo, painéis de enriquecimento, RNA-Seq de célula única, ATAC-seq de célula única, RNA-Seq em massa, e análise de metilação (Tabela 1). São necessárias mais de 30 ferramentas de código aberto para replicar parcialmente a amplitude da funcionalidade no software DRAGEN.[3,4]

Para análise de linha genética, o DRAGEN secondary analysis inclui um conjunto de identificadores de variantes, como ExpansionHunter e identificadores direcionados para genes como *SMN*, *GBA*, *CYP2B6*, *CYP2D6* e *HLA*. O DRAGEN v4.3 também introduz a detecção conjunta em várias regiões (MRJD), um novo identificador especializado que permite a cobertura de genes difíceis em regiões de duplicação segmentar, como *PMS2*, *SMN1*, *SMN2*, *STRC*, *NEB*, *TTN* e *IKBKG*. Essas ferramentas permitem a análise de uma ampla gama de variações genéticas, incluindo variações de nucleotídeo único, inserções e exclusões (indels), expansões de repetição e variações estruturais em regiões genômicas estendidas.

Somente para pesquisa. Não deve ser usado para procedimentos de diagnóstico.

Tabela 1: O DRAGEN secondary analysis é compatível com uma ampla gama de aplicações de análise secundária[a]

Aplicação	Servidor local	Sistemas de sequenciamento Illumina integrados			Plataformas de nuvem da Illumina	
	Servidor do DRAGEN	NovaSeq X Series	Sistemas NextSeq 1000, NextSeq 2000	MiSeq i100 Series	BaseSpace Sequence Hub	Illumina Connected Analytics
Conversão em BCL	✔	✔	✔	✔	✔	Somente personalizado
DRAGEN ORA Compression	✔	✔	✔	✔		Somente personalizado
DRAGEN FASTQ + MultiQC	✔	✔	✔	✔[b]	✔	✔
Genoma completo	Linha genética + somática	Linha genética + somática	Linha genética + somática		Linha genética + somática	Linha genética + somática
Enriquecimento (incluindo exoma)	Linha genética + somática	Linha genética + somática	Linha genética + somática		Linha genética + somática	Linha genética + somática
DRAGEN Amplicon	✔		Somente DNA		✔	✔
RNA	✔	✔	✔		✔	✔
RNA de célula única	✔		✔		✔	✔
NanoString GeoMx NGS			✔		✔	
Metilação	✔	✔			✔	✔
Metagenômica	✔[c]				✔	
COVID, IMAP, IMAP-FLU					✔	
Portfólio TruSight™ Oncology 500	✔				✔[d]	✔
scATAC-Seq	✔				✔	✔
Imputação	✔				✔	✔
Identificador de alelos estrela PGx	✔	✔	✔		✔	✔
Illumina Complete Long Reads					✔	✔
RPIP, RVEK, UPIP, VSP	✔				✔	Beta
Genoma completo pequeno				✔	✔	

a. A versão do software principal DRAGEN varia entre as plataformas, fale com um representante local para obter mais informações.
b. Compatível com o aplicativo de controle de qualidade da biblioteca.
c. Aplicações de metagenômica habilitadas pelo classificador Kmer. Mais ferramentas serão disponibilizadas em breve.
d. Necessária assinatura do Illumina Connected Analytics.
IMAP, Illumina Microbial Amplicon Prep; RPIP, Respiratory Pathogen ID/AMR Panel; RVEK, Respiratory Virus Enrichment Kit; UPIP, Urinary Pathogen ID/AMR Panel; VSP, Viral Surveillance Panel.

Além disso, o mapeamento multigenômico do DRAGEN com referência pangenômica melhora a qualidade do mapeamento, levando a maior precisão na identificação de variantes e resolução de áreas do genoma difíceis de avaliar devido a complexidades de sequência. Isso aumenta a cobertura de genes com potencial relevância clínica e permite a identificação de variantes de nucleotídeo único, de pequenas indels, com variações do número de cópias e variantes estruturais em regiões difíceis de mapear.

Análise eficiente

O software DRAGEN fornece aos laboratórios a velocidade de análise de dados necessária para otimizar a eficiência do processamento de conjuntos de dados NGS. O DRAGEN secondary analysis é executado com array de portas programáveis em campo (FPGA, Field-Programmable Gate Array) acelerada por hardware, garantindo tempos de processamento rápidos. A eficiência dos algoritmos de análise do DRAGEN resultou em dois recordes mundiais de velocidade de análise de dados genômicos.[5,6] Em aplicações práticas, o DRAGEN secondary analysis local pode processar dados de NGS para um genoma completo equivalente a uma cobertura de 40× em cerca de 35 minutos com todos os identificadores[†] em comparação a > 8 horas com métodos de código aberto comumente usados para identificação de um número limitado de tipos de variantes.[7]

Para facilitar o armazenamento, o gerenciamento e o compartilhamento de grandes arquivos de dados de NGS, a tecnologia DRAGEN Original Read Archive (ORA) fornece compactação de até cinco vezes sem perdas de arquivos FASTQ no formato tradicional fastq.gz. A compactação sem perdas do DRAGEN ORA mantém os detalhes dos arquivos FASTQ e é notavelmente rápida, exigindo aproximadamente 8 minutos para a compactação de arquivos FASTQ de 50 a 70 GB, sendo compatível com uma ampla variedade de espécies comumente estudadas. O DRAGEN secondary analysis conta com um conjunto versátil de pipelines que também aceitam arquivos de dados, com arquivos de dados gerados em diferentes estágios dos pipelines (Figura 2).

FPGA e aceleração de hardware

A FPGA altamente configurável permite implementações aceleradas por hardware ultraeficientes de algoritmos de análise genômica, como conversão de arquivos de identificação de bases (BCL), mapeamento, alinhamento, classificação, marcação duplicada e identificação de variantes de haplótipos. A natureza flexível das FPGAs permite que a Illumina desenvolva um amplo pacote de pipelines de aplicações do DRAGEN, com atualizações e adições frequentes para fornecer a melhor precisão, abrangência e eficiência possíveis.

† Com base nos dados internos da Illumina de acordo com os padrões HG001-HG007 no servidor DRAGEN v4, sem novos identificadores especializados, como o MRJD e repetições em tandem de número variável, disponíveis no DRAGEN v4.3.

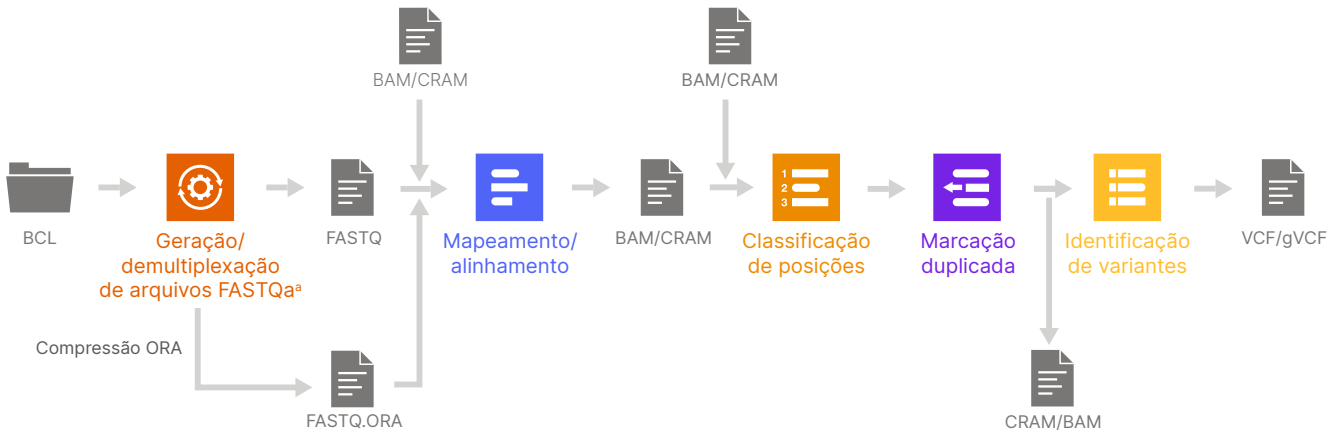

Figura 2: Flexibilidade de pipelines do DRAGEN secondary analysis: cada pipeline do DRAGEN contém um conjunto específico de tapas para oferecer uma análise precisa e eficiente. O pipeline de exemplo da linha genética do genoma completo do DRAGEN oferece a flexibilidade de aceitar vários arquivos de entrada e produzir uma variedade de tipos de saída, permitindo que os usuários personalizem sua experiência e produzam o formato de arquivo desejado.

a. A conversão em BCL também está disponível como uma ferramenta independente.

Somente para pesquisa. Não deve ser usado para procedimentos de diagnóstico.

Referências personalizadas

O DRAGEN secondary analysis permite que os usuários gerem uma referência personalizada humana, não humana ou não padrão. As referências criadas podem ser usadas como entrada para todas as aplicações do DRAGEN compatíveis com arquivos de referência personalizados. A maioria dos pipelines do DRAGEN inclui suporte integrado para hg19, hg38 (com ou sem HLA), GRCh37, CHM13v2 e hs37d5. O software DRAGEN permite que os usuários ampliem os recursos de mapeamento multigênico padrão de gráficos para populações diversas e específicas.

Escalabilidade

O DRAGEN secondary analysis possibilita que os laboratórios dimensionem as operações conforme necessário, mantendo os custos e os tempos de resposta baixos. O software DRAGEN pode facilitar a expansão dos recursos de pesquisa de várias maneiras:

1. **Mantendo-se atualizado com o NovaSeq™ X Series:** o DRAGEN integrado pode processar até quatro aplicativos simultaneamente por lâmina de fluxo em uma corrida única, incluindo no máximo uma conversão em BCL e três outros pipelines à sua escolha.

2. **Capacidade de disparo:** em situações de aumento de cargas de trabalho com altos volumes de amostras, os laboratórios podem aproveitar a capacidade adicional na nuvem com o DRAGEN secondary analysis nos aplicativos Illumina Connected Analytics ou DRAGEN no BaseSpace Sequence Hub (Figura 3).

3. **Expansão das operações:** uma única instância do DRAGEN pode executar uma ampla gama de pipelines do DRAGEN e tipos de amostras compatíveis. A abrangência e a eficiência do software DRAGEN permitem que os usuários aumentem as operações sem comprometer os tempos de resposta ou a qualidade dos resultados.

4. **Transição para genomas:** os pipelines pré-construídos do DRAGEN possibilitam a transição fácil entre painéis direcionados de exomas para genomas.

5. **Iniciativas genômicas de grandes populações:** o DRAGEN secondary analysis oferece um fluxo de trabalho simplificado para análise de coortes de grande escala, apresentando vários pipelines usados em conjunto para identificar variações genéticas com alta precisão. O DRAGEN gVCF Genotyper possibilita agregar de milhares a milhões de arquivos no formato de identificação de variante genômica (gVCF) e incorpora novos lotes sem reprocessar lotes existentes. A compactação ORA economiza custos de armazenamento.

6. **Aplicações de sequenciamento profundo:** o DRAGEN secondary analysis é compatível com a análise de sequenciamento de dados de alta profundidade com alta eficiência para a cobertura média de mais de 300 vezes para genomas e 1.000 vezes para exomas. Os recursos de sequenciamento profundo são valiosos para aplicações como pesquisa oncológica e estudos de doenças genéticas raras.

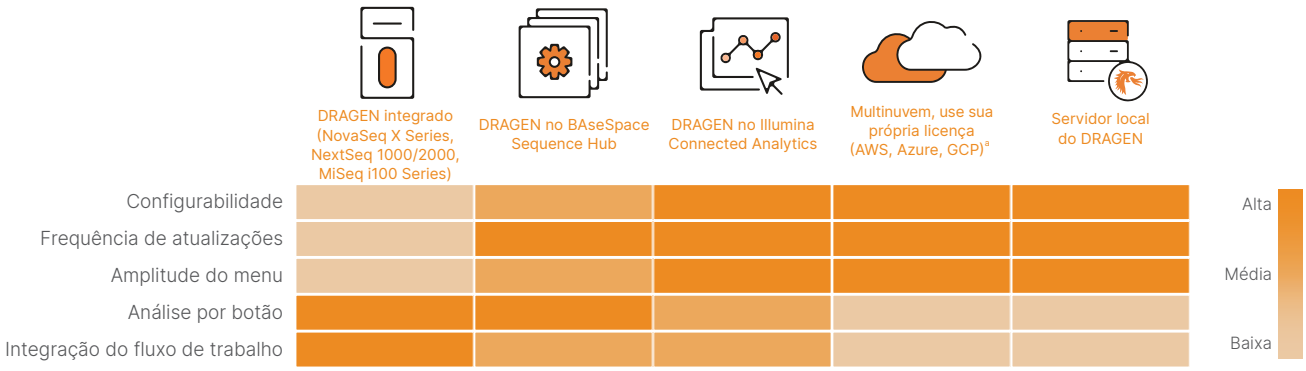

Figura 3: Opções de acesso ao pipeline do DRAGEN com recursos projetados para atender às necessidades de análise de NGS de cada laboratório.

a. Entre em contato com seu representante da Illumina para obter informações de acesso a Amazon Web Service (AWS), Azure ou Google Cloud Platform (GCP, acesso antecipado).

Acessibilidade multiplataforma

O pacote de pipelines do DRAGEN pode ser acessado por meio de soluções disponíveis no local, no instrumento ou na nuvem, permitindo que os laboratórios selecionem a solução que melhor atenda às suas necessidades (Figura 3).

Servidor local do DRAGEN

O servidor local do DRAGEN conta com uma solução de armazenamento local para coletar e armazenar dados de NGS. As especificações do DRAGEN Server v4 estão listadas na Tabela 2. Após a transferência dos dados brutos do instrumento de sequenciamento para o armazenamento local por meio da conexão de rede local, eles são acessados pelo servidor do DRAGEN para realizar o fluxo de trabalho selecionado. Após a análise, o software grava os arquivos de dados gerados no armazenamento local. Servidor local do DRAGEN:

- É compatível com configuração flexível de recursos do DRAGEN por meio de uma interface de linha de comando
- Substitui até 30 instâncias de computação tradicionais
- Processa dados de NGS para um genoma humano inteiro com cobertura de 40X em aproximadamente 35 minutos
- Compatível com o Illumina Connected Insights-Local, fornecendo interpretação de variantes e geração de relatórios para pesquisa clínica oncológica

Tabela 2: Especificações do DRAGEN Server v4

Componente	Especificações
CPU	Dual Intel Xeon Gold 6226R 2.9 GHx, 16C / 32T
Memória do sistema	DDR4 de 512 GB
Scratch drive	2 × NVMe de 7,68 TB
Unidade do SO	2× SSD de 480 GB (RAID 1)
Aceleração de hardware	1× placa FPGA
Fator de forma	2U
Dimensões	A 8,8 cm (3,5 pol.), L 43,8 cm (17,2 pol.), D 76,4 cm (29,9 pol.)
Fonte de alimentação	Fonte de alimentação redundante, dupla, Hotswap de 1.968 W

DRAGEN integrado ao NovaSeq X Series

O NovaSeq X Series inclui o DRAGEN secondary analysis integrado, oferecendo análise precisa, automatizada e simplificada, projetada para lidar com o volume extraordinário de dados gerados pelo NovaSeq X Series. O pacote de software DRAGEN integrado fornece análise secundária e compactação ORA com aplicações de NGS (Tabela 1), abrangendo conversão em BCL, linha genética, somática, enriquecimento, RNA e metilação. O DRAGEN integrado:

- Executa vários pipelines de análise secundária em paralelo
- Em uma corrida única, executa até quatro aplicações diferentes por lâmina de fluxo simultaneamente
- Oferece compactação de dados de até 5x sem perdas, com economia de custos de armazenamento
- Proporciona economia em análises que pode exceder o preço de compra do NovaSeq X Series em cinco anos

Sistemas NextSeq 1000 e NextSeq 2000 com DRAGEN integrado

Os sistemas NextSeq 1000 e NextSeq 2000 incluem o software DRAGEN integrado para realização de análise secundária rápida e precisa. O software é acessado por meio de uma interface gráfica fácil de usar, permitindo que usuários, especialistas ou não, realizem as análises necessárias e produzam resultados rapidamente. O software DRAGEN integrado oferece um conjunto selecionado de pipelines para diversas aplicações comuns de NGS (Tabela 1) e inclui a premiada análise de ML e mapeamento multigenômico para uma identificação de variantes de alta qualidade. O DRAGEN integrado:

- Oferece a mais alta precisão para qualquer sistema de sequenciamento de bancada com DRAGEN secondary analysis integrado
- Fornece acesso aos pipelines selecionados de informática do DRAGEN
- Possibilita que usuários gerem resultados em apenas duas horas
- Usa algoritmos intuitivos de pipeline para reduzir a dependência de especialistas em informática externos

Somente para pesquisa. Não deve ser usado para procedimentos de diagnóstico.

DRAGEN integrado ao MiSeq i100 Series

O MiSeq i100 Series fornece análise intuitiva ultrarrápida, incluindo conversão em BCL DRAGEN, controle de qualidade da biblioteca DRAGEN, sequenciamento de genoma pequeno completo DRAGEN e DRAGEN Microbial Enrichment Plus.

- Resultados rápidos com análise secundária abrangente gerada em duas horas ou menos‡
- Fluxo de trabalho altamente eficiente com um único ponto de contato do usuário para VCF e/ou relatório e sem transferências intermediárias de arquivos
- Excepcionalmente fácil com uma interface intuitiva para usuários não especializados

BaseSpace Sequence Hub

O pacote DRAGEN baseado na nuvem disponível no BaseSpace Sequence Hub combina a análise precisa e eficiente com um ecossistema seguro e uma escalabilidade versátil. O software DRAGEN no BaseSpace Sequence Hub permite uma análise secundária por botão para laboratórios de todos os tamanhos e disciplinas. O BaseSpace Sequence Hub é uma extensão direta dos instrumentos da Illumina. O fluxo de dados criptografados do instrumento para o BaseSpace Sequence Hub permite o gerenciamento e a análise de dados com facilidade usando um conjunto selecionado de aplicações em um ambiente seguro. O BaseSpace Sequence Hub, desenvolvido pela Amazon Web Services (AWS):

- Oferece uma solução simples e fácil de usar para análise do DRAGEN
- Usa uma interface gráfica intuitiva para operação eficiente por usuários especialistas ou não
- Fornece acesso a recursos de computação poderosos sem gastos de capital para infraestrutura adicional

Illumina Connected Analytics

Illumina Connected Analytics é uma plataforma abrangente de bioinformática baseada na nuvem que capacita os pesquisadores para gerenciamento, análise e interpretação de grandes volumes de dados multiômicos em um ambiente seguro, escalável e flexível. Acesse o pacote do DRAGEN secondary analysis no Illumina Connected Analytics, disponível como pacote de pipelines pré-definidos ou ferramentas individuais para incorporação em pipelines personalizados.

‡ Quando executado de acordo com as recomendações da amostra.

Resumo

O DRAGEN secondary analysis é um poderoso pacote de ferramentas de software que fornece análise precisa, abrangente e eficiente dos dados de NGS. Várias opções de implantação do software DRAGEN possibilitam que os laboratórios escolham a solução que melhor se adapte ao tipo e à escala de seus projetos. Além disso, os usuários podem combinar várias opções de implantação para melhor atender às suas necessidades de desempenho e fluxo de trabalho. À medida que a tecnologia de NGS continua a progredir, atualizações oportunas do DRAGEN secondary analysis garantem o melhor desempenho possível dos pipelines atuais, enquanto novos pipelines continuam a ser adicionados conforme as aplicações são disponibilizadas.

Saiba mais

DRAGEN secondary analysis

Página de suporte do DRAGEN secondary analysis

Contate-nos

Referências

1. US Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10/results. Acessado em 14 de março de 2022.
2. Illumina. DRAGEN sets new standard for data accuracy in PrecisionFDA benchmark data. Optimizing variant calling performance with Illumina machine learning and DRAGEN graph. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Publicado em 12 de janeiro de 2022. Acessado em 11 de julho de 2024.
3. Illumina. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Publicado em 9 de novembro de 2020. Acessado em 11 de julho de 2024.
4. Dados internos em arquivo. Illumina, Inc., 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Publicado em 23 de outubro de 2017. Acessado em 11 de julho de 2024.
6. The San Diego Union-Tribune. Rady Children's Institute sets Guinness world record. sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/. Publicado em 12 de fevereiro de 2018. Acessado em 11 de julho de 2024.
7. Betschart RO, Thiéry A, Aguilera-Garcia D, et al. Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment. *Sci Rep.* 2022;12(1):21502. doi:10.1038/s41598-022-26181-3.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Publicado em 24 de maio de 2021. Acessado em 14 de março de 2022.

+1 (800) 809-4566, ligação gratuita (EUA) | tel. +1 (858) 202-4566
techsupport@illumina.com | www.illumina.com

© 2025 Illumina, Inc. Todos os direitos reservados. Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.
M-GL-00680 PTB v11.0