

# DRAGEN<sup>™</sup> secondary analysis

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

**illumina**<sup>®</sup>

## 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 Illumina DRAGEN (Dynamic Read Analysis for GENomics, análise de leitura dinâmica para genômica) secondary analysis foi desenvolvido para abordar desafios importantes associados à análise de dados de NGS relacionada a uma ampla variedade de aplicações, incluindo genoma completo, exoma, transcriptoma, estudos 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. No 2020 Precision FDA Truth Challenge V2 (PrecisionFDA V2), o DRAGEN secondary analysis v3.7 venceu como o mais preciso em todas as regiões de referência e de difícil mapeamento, com dados de sequenciamento da Illumina.<sup>1,2</sup> As versões subsequentes continuam a estabelecer novos padrões de precisão, com avanços em áreas que incluem a tecnologia de aprendizado de máquina (ML) e o DRAGEN multigenome (gráfico). A versão mais recente do DRAGEN secondary analysis v4.3 fornece precisão sem precedentes na identificação de variantes pequenas com pontuação F1 de 99,89% (medida combinada de precisão e recall) em todas as regiões de referência (figura 1). Isso é possibilitado pela referência do DRAGEN multigenome (gráfico) de terceira geração baseada em 128 amostras com 256 haplótipos de dados de referência pangenômicos integrados, capturando uma maior diversidade genética. O novo 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%.

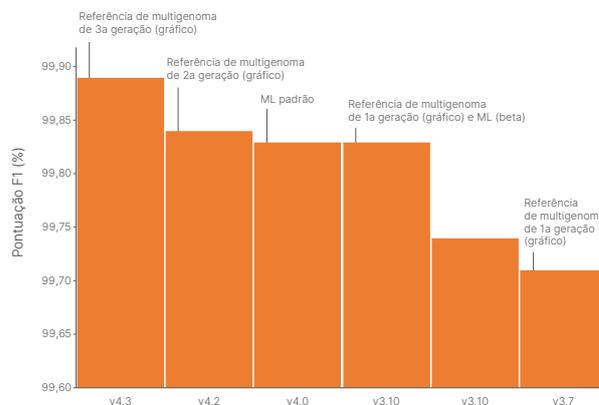


Figura 1: Precisão do DRAGEN secondary analysis: a pontuação F1 do eixo Y (%) é um cálculo de resultados positivos e negativos verdadeiros como uma proporção dos resultados totais.<sup>3,4</sup>

## 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 o sequenciamento do genoma completo (WGS), o 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.<sup>3,4</sup>

Para análise de linha genética, o DRAGEN secondary analysis inclui um pacote de identificadores de variantes, como o ExpansionHunter e identificadores direcionados, como *SMN*, *GBA*, *CYP2B6*, *CYP2D6* e *HLA*. O DRAGEN v4.3 também conta com o 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 *IKBK*. 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. Além disso, a referência do DRAGEN multigenome (gráfico) melhora a qualidade do mapeamento, levando a uma maior precisão na identificação de variantes e resolvendo áreas do genoma difíceis de avaliar devido às complexidades da 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.

Tabela 1: O DRAGEN secondary analysis é compatível com uma ampla gama de aplicações de análise secundária<sup>a</sup>

Aplicação	Servidor local	Sistemas de sequenciamento Illumina integrados		Plataformas de nuvem da Illumina	
	Servidor do DRAGEN	NovaSeq X Series	NextSeq 1000, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics
Conversão em BCL	✓	✓	✓	✓	Somente personalizado
DRAGEN ORA Compression	✓	✓	✓		Somente personalizado
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
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	✓ <sup>b</sup>			✓	
RNA pathogen detection				✓	
COVID	COVIDSeq. Linhagem da COVID		COVIDSeq (somente nuvem)	COVIDSeq. Linhagem da COVID	
Portfólio TruSight™ Oncology 500	✓			✓ <sup>c</sup>	✓
scATAC-Seq	✓			✓	✓
Imputação	✓			✓	✓
Identificador de alelos estrela PGx	✓	✓	✓	✓	✓
Illumina Complete Long Reads				✓	✓
DRAGEN secondary analysis para RPIP e UPIP	✓			✓	Beta

a. A versão do software principal DRAGEN varia entre as plataformas, fale com um representante local para obter mais informações.

b. Aplicações de metagenômica habilitadas pelo classificador Kmer. Mais ferramentas serão disponibilizadas em breve.

c. Necessária assinatura do Illumina Connected Analytics.

## Análise eficiente

O software DRAGEN foi desenvolvido para conceder aos laboratórios a velocidade de análise de dados necessária para otimizar a eficiência do processamento de conjuntos de dados de NGS. O DRAGEN secondary analysis é acelerado por hardware e usa arquitetura de matriz de portas programáveis em campo (FPGA) para alcançar tempos de resposta rápidos. A eficiência dos algoritmos de análise do DRAGEN resultou em dois recordes mundiais de velocidade de análise de dados genômicos.<sup>5,6</sup> 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<sup>7</sup> em comparação a > 8 horas com métodos de código aberto comumente usados para identificação de um número limitado de tipos de variantes.<sup>7</sup>

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.

\* 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 o VNTR, disponíveis no DRAGEN v4.3.

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 de entrada e criam arquivos de saída 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 variante de haplótipo. 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.

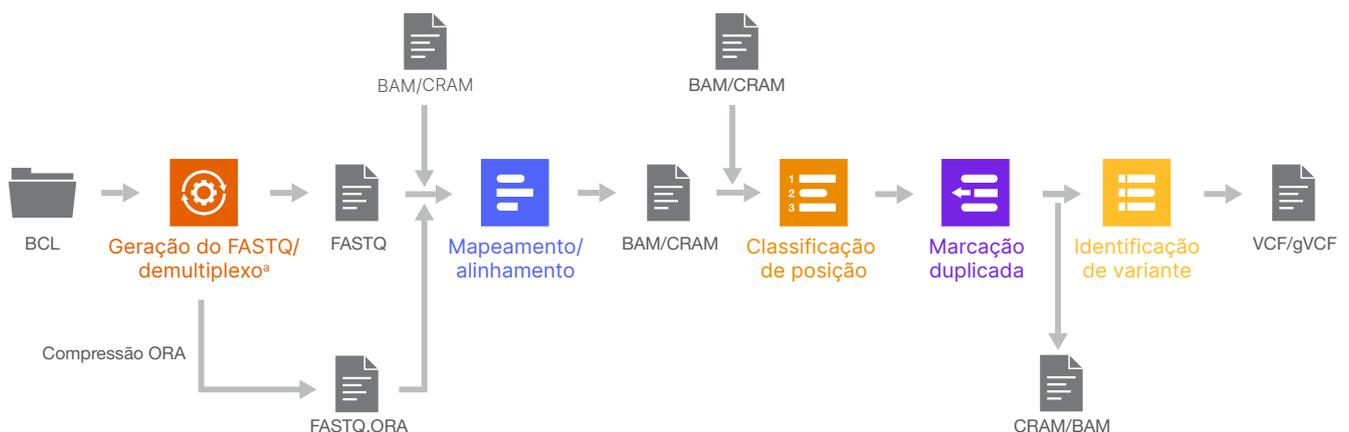


Figura 2: Flexibilidade de pipelines do DRAGEN secondary analysis: cada pipeline do DRAGEN contém um conjunto específico de etapas para oferecer uma análise precisa e eficiente. O pipeline de exemplo da linha genética do genoma completo do DRAGEN oferece a flexibilidade de aceitar vários arquivos de 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.

## 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 referência multigênica 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. Manter-se atualizado com o NovaSeq™ X Series e os sistemas NextSeq™ 1000 e NextSeq 2000:** o DRAGEN integrado pode executar várias aplicações simultâneas (quatro aplicações simultâneas com o máximo de uma conversão em BCL e três outros pipelines de sua escolha) por lâmina de fluxo em uma execução única.
- 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 alta profundidade com alta eficiência para a cobertura média de mais de 300 vezes para genomas e 1000 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. Após a transferência dos dados brutos do instrumento de sequenciamento para o armazenamento local por meio da conexão de rede local, eles são movidos para o servidor do DRAGEN para realizar o fluxo de trabalho selecionado. Após a análise, o software grava os arquivos de saída gerados de volta no local de 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.

## 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 execução 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 System 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 seletivo de pipelines para cobrir uma variedade de aplicações de NGS comuns (tabela 1) e inclui análises de referência de ML e de multigenoma (gráfico) premiadas para 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.

## 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.

## 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](#)



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

© 2024 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](http://www.illumina.com/company/legal.html).  
M-GL-00680 PTB v10.0

## Referências

1. The Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. <https://precision.fda.gov/challenges/10/results>. Acessado em 14 de março de 2022.
2. Catreux S, Jain V, Murray L, et al. DRAGEN sets new standard for data accuracy in PrecisionFDA benchmark data. Optimizing variant calling performance with Illumina machine learning and DRAGEN graph. Site da Illumina. <https://www.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. Mehio R, Ruehle M, Catreux S, et al. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. Site da Illumina. [illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html](https://www.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](https://www.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. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.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. Publicado em 13 de dezembro de 2022. doi:10.1038/s41598-022-26181-3
8. Gross A, Maciucă S, Cox A, et al. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. Site da Illumina. [www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html](https://www.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.