

# Análise de CNV de linha genética através do fluxo de trabalho do Illumina DNA Prep with Exome 2.5 Enrichment

Detecção eficiente de variantes  
do número de cópias com um  
painel de normais



## Introdução

As variações do número de cópias (CNVs) são uma fonte importante de diversidade genética em humanos.<sup>1</sup> Essa classe de variação genômica ocorre quando o número de cópias de um gene ou região genômica varia entre indivíduos e inclui grandes duplicações ou deleções genômicas. A análise de CNV pode ajudar os pesquisadores a identificar variantes associadas a traços complexos ou suscetibilidade a doenças, como câncer, doenças autoimunes, distúrbios genéticos hereditários e muito mais.<sup>1-4</sup>

Em comparação com o uso de microarrays para detecção de CNV, os métodos de sequenciamento de última geração (NGS) revelam mais detalhes sobre variações estruturais. Além disso, o NGS pode detectar algumas variantes inferiores a 50 kb, que não são detectadas pelos arrays. O Illumina DNA Prep with Exome 2.5 Enrichment faz parte de uma solução de alto desempenho, rápida e confiável de sequenciamento do exoma completo (WES) de seres humanos, gerando bibliotecas compatíveis com a análise de CNV do NGS. Combinado com os sistemas de sequenciamento e pipelines de análise da Illumina, este kit de preparação de biblioteca pode ser usado para detectar variantes de nucleotídeo único (SNV), pequenas inserções e deleções (indels) e grandes CNVs estruturais.

Esta nota de aplicação demonstra um fluxo de trabalho simplificado de sequenciamento de exoma para análise de CNV da linha genética, desde a preparação da biblioteca até a obtenção de insights (Figura 1). A solução de WES integra o Illumina DNA Prep with Exome 2.5 Enrichment, o qualificado Illumina NGS e o altamente preciso DRAGEN™ secondary analysis para a identificação de CNVs.

Os usuários também podem utilizar a plataforma de pesquisa de interpretação de variantes Emedgene™, com tecnologia de inteligência artificial explicável (XAI). O software Emedgene possibilita a análise de variantes e a automação de insights, da curadoria e do procedimento operacional padrão (POP), além de incluir opções simplificadas para a geração de relatórios de pesquisa.

## Métodos

### Amostras

Para viabilizar a detecção de CNV de linha genética no exoma, os usuários devem desenvolver um arquivo de referência de linha de base a partir de um painel de amostras supostamente normais para comparação durante a análise. Esse painel de referência de normais é usado para normalizar as contagens-alvo e permite a detecção precisa de CNVs durante a avaliação de amostras de casos. As amostras do painel de normais devem ser preparadas e sequenciadas nas mesmas condições que as amostras do caso.

Para criar o painel de normais, 54 amostras (Coriell Institute for Medical Research) sem CNVs patogênicos conhecidos (Tabela 1) foram selecionadas para representar a diversidade de variação natural em superpopulações humanas (com base no 1000 Genomes Project, Tabela 2).<sup>5</sup> Para demonstrar o desempenho do fluxo de trabalho de WES na detecção de CNV, 17 amostras de casos contendo um CNV detectado anteriormente dentro de um gene ou região de interesse foram usadas em uma análise com o painel de normais.



Figura 1: Fluxo de trabalho de WES abrangente para identificação de CNV. Combinação do Illumina DNA Prep with Exome 2.5 Enrichment, das plataformas Illumina NGS, da análise rápida de dados com o DRAGEN secondary analysis e da interpretação de variantes com o software Emedgene para obtenção de um fluxo de trabalho fácil de usar e simplificado, a fim de realizar um sequenciamento preciso do exoma para interrogação de CNVs.

Tabela 1: IDs de amostras do Coriell Institute incluídas no painel de normais.

HG00096	HG01393	HG01950	HG03006	NA10830	NA12877	NA18502	NA18970	NA21112
HG00190	HG01441	HG01985	HG03870	NA10831	NA12878	NA18508	NA19681	NA24143
HG00262	HG01551	HG01990	HG03882	NA10835	NA12889	NA18622	NA19720	NA24149
HG00626	HG01599	HG02013	HG03898	NA10838	NA12890	NA18637	NA20509	NA24631
HG00628	HG01896	HG02348	HG04090	NA10839	NA12891	NA18942	NA20875	NA24694
HG01392	HG01914	HG02521	HG04214	NA12249	NA12892	NA18957	NA21098	NA24695

## Preparação da biblioteca

As bibliotecas de sequenciamento do exoma foram preparadas em duplicata, seguindo o protocolo do Illumina DNA Prep with Exome 2.5 Enrichment (documento n.º 1000000048041 v07) com 50 ng de gDNA de amostras do Coriell Institute, salvo indicação em contrário. Os produtos IDT for Illumina DNA/RNA UD Indexes Sets A-D, Tagmentation (Illumina, n.ºs de catálogo 0027213, 20027214, 20042666 e 20042667) foram usados para permitir a indexação exclusiva em todas as bibliotecas preparadas. As bibliotecas de pré-enriquecimento foram agrupadas por massa (alvo de 250 ng por amostra) a 12-plex por pool e hibridizadas de um dia para o outro usando o Twist Bioscience for Illumina Exome 2.5 Panel, disponível como parte dos kits Illumina DNA Prep with Exome 2.5 Enrichment (Illumina, n.ºs de catálogo 20077595 e 20077596). As bibliotecas foram preparadas com e sem o spike-in do Twist Bioscience for Illumina Mitochondrial Panel (Illumina, n.º de catálogo 20093180). O spike-in do painel mitocondrial seguiu a diluição 1:100 descrita no Guia de referência do Illumina DNA Prep with Exome 2.5 Enrichment (documento n.º 1000000157112).

## Sequenciamento

As bibliotecas enriquecidas foram sequenciadas nos sistemas NovaSeq™ 6000 e NextSeq™ 2000 usando reagentes de sequenciamento por síntese (SBS) padrão e sequenciamentos tipo paired-end de 150 bp. As saídas mais altas de dados disponíveis foram usadas para alcançar uma alta profundidade de cobertura por amostra e possibilitar a redução da amostragem como parte da análise. O painel de amostras normais e as amostras do caso foram sequenciadas com o NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, n.º de catálogo 20028312) e o NextSeq 2000 P3 Reagents (300 cycles) (Illumina, n.º de catálogo 20040561). Os kits de reagentes do NovaSeq 6000 System foram projetados para facilitar o uso com três cartuchos prontos para uso pré-carregados com todos os reagentes necessários.

Tabela 2: Atributos resumidos no painel de normais.

Sexo	N.º de amostras
Masculino	27
Feminino	27
Superpopulação <sup>5</sup>	
Europeu (EUR)	18
Americano mestiço (AMR)	8
Africano (AFR)	7
Sul-asiático (SAS)	9
Asiático do Leste (EAS)	12

Tabela 3: Rendimento e cobertura de sequenciamento entre plataformas relacionados ao painel de normais e às amostras de CNV.

Plataforma	Amostras por sequenciamento	Cobertura média por amostra
NovaSeq 6000 System	72	Aprox. 540×
NextSeq 2000 System	12	Aprox. 280×

O NextSeq 2000 System fornece flexibilidade com várias configurações de lâmina de fluxo e oferece a análise do DRAGEN integrada. O número de amostras por sequenciamento e a cobertura média alcançada por amostra em todas os sequenciamentos são mostrados na [Tabela 3](#).

## Análise de dados

O aplicativo DRAGEN Baseline Builder v4.2, acessível no BaseSpace™ Sequence Hub, foi usado para criar o painel de arquivos de contagens de alvos individuais normais com o genoma de referência não gráfico hg38. Os dados de sequenciamento para amostras de casos de CNV tiveram a amostra reduzida para 80 milhões de leituras (cerca de 120 a 140× a profundidade média de cobertura no alvo) e 50 milhões de leituras (cerca de 70 a 90× a profundidade média de cobertura no alvo) e foram analisados usando o aplicativo DRAGEN Enrichment v4.2, também acessível no BaseSpace Sequence Hub, com detecção de CNV ativada. Para cada plataforma, o painel correspondente de normais foi usado para normalização de amostras. O aplicativo DRAGEN Enrichment realiza análises secundárias precisas e eficientes para identificação abrangente de variantes, incluindo SNV, indels, CNV e variantes estruturais (SV), entre outras aplicações. A identificação da variante do DRAGEN para o Illumina DNA Prep with Exome 2.5 Enrichment também pode ser realizada no NextSeq 2000 System ou pode ser totalmente integrada ao fluxo de trabalho de pesquisa de interpretação de variante do software Emedgene.

## Resultados

### Detecção de CNV em muitas variações

A análise com o pipeline DRAGEN CNV no aplicativo DRAGEN Enrichment revelou uma detecção robusta de CNV para as variantes avaliadas neste estudo em apenas 50 milhões de leituras, ou profundidade média de cobertura de aproximadamente 70 a 90 vezes, em todas as plataformas. A detecção de CNVs é determinada pela fórmula: (número de regiões identificadas)/(regiões que se sobrepõem ao painel do exoma e à verdade conhecida). As CNVs eram consideradas detectadas quando essa métrica era > 98% (Tabela 4).

Muitas das variantes testadas são encontradas no gene *DMD*, que é um alvo importante na pesquisa de doenças genéticas (Tabela 4).<sup>6</sup> Com o Integrated Genomics Viewer, os eventos esperados de CNV no *DMD* detectados podem ser visualizados nas sete amostras afetadas testadas (Figura 2). Isso demonstra a capacidade de detecção de CNVs nesse gene ao usar o fluxo de trabalho do Illumina Exome 2.5. O rastreo de arquivos BED do painel do Exome 2.5 mostra as regiões cobertas pelo painel (Figura 2, parte superior). Os resultados dependerão de vários fatores, e é recomendado sequenciar a cobertura média acima de 150 vezes para testes iniciais de um fluxo de trabalho de laboratório a fim de garantir cobertura suficiente para a maioria das variantes.

### Possibilidade de análise adicional com o software Emedgene

O software Emedgene simplifica e dimensiona a interpretação de variantes, economizando de 50% a 75% do tempo por caso de pesquisa.<sup>7</sup> Vários recursos potencializam a interpretação definida pelo usuário, incluindo a tecnologia de XAI para classificações automatizadas transparentes e baseadas em evidências de variantes potencialmente causais para amostras; um gráfico de anotações e conhecimentos sempre atualizado; a visualização de variantes; a curadoria de variantes; a automação definida pelo usuário e muito mais, promovendo uma interpretação eficiente e informada das variantes. O software Emedgene foi projetado para oferecer uma experiência de usuário eficiente e intuitiva.

Para simplificar a análise de casos, a priorização de variantes de IA ou a funcionalidade de "lista de opções" compila variantes, incluindo CNVs, com maior probabilidade de resolução de um caso. Essa funcionalidade inclui evidências de apoio e proporciona economias significativas de tempo para o analista. Em um estudo separado<sup>8</sup> de 51 singletons previamente resolvidos por uma variante de CNV, a variante de resolução foi identificada em uma lista de 10 variantes em 92% dos casos. Em 6% (n = 3), a variante de resolução estava presente na lista de candidatas. Durante a revisão do caso, o software Emedgene permite desmarcar variantes selecionadas pela lista ou marcar manualmente variantes não selecionadas por ela. A classificação automatizada do ACMG\* também abrange CNVs.

### Painel de acesso de arquivos normais

O painel de arquivos normais é gerado com amostras do Coriell Institute e fornece aos pesquisadores um conjunto de dados normais conhecidos para comparação com amostras de casos sem a necessidade de gastar tempo e recursos adquirindo, sequenciando e analisando esse conjunto de amostras. Esses arquivos padrão são usados de forma mais eficiente em testes iniciais do fluxo de trabalho do Illumina DNA Prep with Exome 2.5 Enrichment para análise de CNV. Recomenda-se que os laboratórios desenvolvam um painel de normais específicos para seus protocolos laboratoriais. Se estiverem usando esses arquivos para testes iniciais, as bibliotecas devem ser preparadas para amostras de casos seguindo o mesmo protocolo usado para preparar o painel de normais. Desvios no protocolo podem levar a diferenças nos perfis de cobertura, reduzindo a eficácia do painel padrão de normais. Arquivos e informações adicionais podem ser encontrados no site de suporte da Illumina.

\*ACMG, American College of Medical Genetics and Genomics.

Tabela 4: Detecção de CNV classificada por tamanho aproximado do evento com base nas coordenadas de evento esperadas.

ID do Coriell	Cromossomo	Localização genética (éxons afetados) ou cromossômica	Evento esperado	Tamanho aproximado do evento (kb) <sup>a</sup>	Detectado em 50 milhões de leituras?
NA04315	X	<i>DMD</i> (44)	Perda	0,15	✓
NA05115	X	<i>DMD</i> (45)	Perda	0,18	✓
NA05117	X	<i>DMD</i> (45)	Perda	0,18	✓
NA23599	X	<i>MECP2</i> (3–4)	Perda	2,25	✓
NA18949	17	<i>BRCA1</i> (15–16)	Perda	3,59	✓
NA21939	15	<i>FBN1</i> (42–43)	Perda	3,70	✓
HG03857	16	<i>PALB2</i> (5–7)	Perda	4,23	✓
HG00343	22	<i>CHEK2</i> (9–10)	Perda	4,25	✓
NA23127	X	<i>DMD</i> (27–28)	Ganho	7,47	✓
NA04520	16	<i>TSC2</i> (1–15)	Perda	16,28	✓
NA04327	X	<i>DMD</i> (5–7)	Ganho	22,70	✓
NA10283	X	<i>DMD</i> (72–79)	Perda	54,38	✓
HG00500	2	<i>SPAST</i> (4–17)	Ganho	58,84	✓
NA23675	X	<i>MECP2</i> (todos)	Ganho	76,14	✓
NA23087	X	<i>DMD</i> (2–30)	Ganho	608,45	✓
NA06870	18	18p11.32-18p11.1	Ganho	15.390,21	✓
NA20125	10	10q23.1-10q26.3	Ganho	52.877,99	✓

a. Com base nas coordenadas de eventos esperados do éxon afetado ou da localização cromossômica.

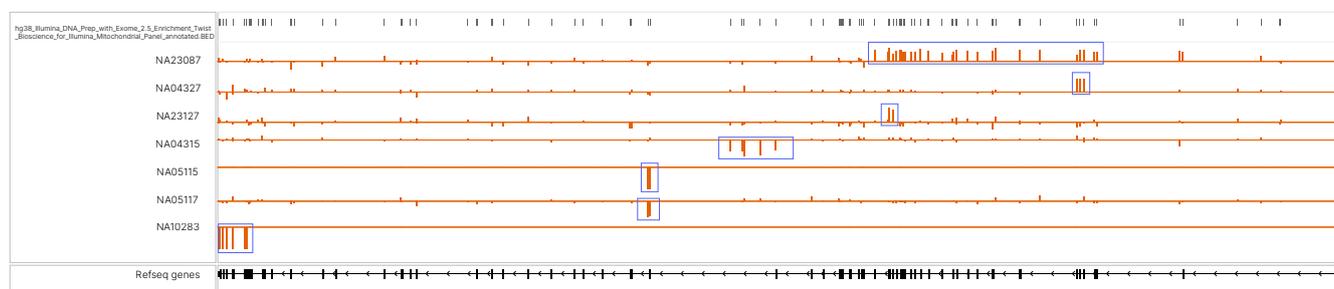


Figura 2: Visualização de CNVs detectadas em sete amostras que contêm eventos esperados no gene *DMD*. Arquivos que fornecem uma representação BigWig do sinal normalizado tangente (\*.tn.bw) são produzidos como parte da análise do aplicativo DRAGEN Enrichment e mostram ganhos e perdas com base nas regiões-alvo no painel de enriquecimento Twist Bioscience for Illumina Exome 2.5. Os rastreios são mostrados com escalonamento automático e como um gráfico de barras.

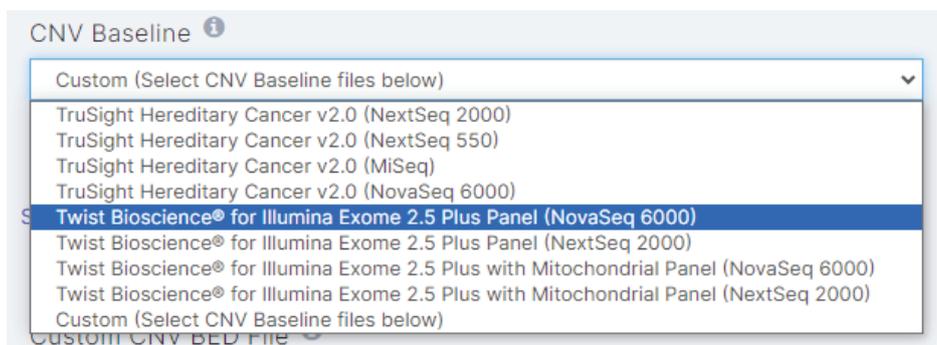


Figura 3: Menu suspenso do painel de normais no aplicativo DRAGEN Enrichment v4.3 no BaseSpace Sequence Hub.

Ao usar o BaseSpace Sequence Hub, o aplicativo DRAGEN Enrichment v4.3 fornece opções no menu suspenso a fim de realizar a análise inicial de CNV com o painel de enriquecimento Twist Bioscience for Illumina Exome 2.5 para várias referências (hg19, hg38 e hs37d5 gráficos e não gráficos) em várias plataformas (NextSeq 2000 e NovaSeq 6000 Systems) usando várias bibliotecas (Painel Twist Bioscience for Illumina Exome 2.5 Enrichment com e sem Twist Bioscience for Illumina Mitochondrial Panel spike-in) (Figura 3). As amostras de caso usadas com este fluxo de trabalho de análise devem ser preparadas conforme descrito na seção Métodos desta nota de aplicação.

## Resumo

O fluxo de trabalho do exoma do NGS para análise de CNVs usa o Illumina DNA Prep with Exome 2.5 Enrichment para preparação de bibliotecas, sequenciamento nos sistemas NovaSeq 6000 ou NextSeq 2000 e aplicativos DRAGEN para análise de dados. Os resultados demonstram a aplicação de um painel de normais como referência para a análise de CNV de amostras de caso através do fluxo de trabalho de WES descrito. A alta identificação de CNVs pelo software DRAGEN secondary analysis se correlaciona bem com outros métodos. Esse fluxo de trabalho, que é usado na análise de CNV, incluindo o uso do painel de normais, também pode ser adaptado a outras plataformas de sequenciamento da Illumina. O software Emedgene ajuda os laboratórios a realizarem análises adicionais, como interpretação de variantes e geração de relatórios de pesquisa.

## Saiba mais

- [Painel de dados de referência de normais](#)
- [Análise da variante do número de cópias](#)
- [Illumina DNA Prep with Exome 2.5 Enrichment](#)
- [Plataformas de sequenciamento da Illumina](#)
- [DRAGEN secondary analysis](#)
- [Aplicativo DRAGEN Germline](#)
- [Software Emedgene](#)

## Referências

1. Zhang F, Gu W, Hurles ME, Lupski JR. [Copy number variation in human health, disease, and evolution](#). *Annu Rev Genomics Hum Genet.* 2009;10:451-481. doi:10.1146/annurev.genom.9.081307.164217
2. Mehawej C, Maalouf JE, Abdelkhalik M, Mahfouz P, Chouery E, Megarbane A. [CNV Analysis through Exome Sequencing Reveals a Large Duplication Involved in Sex Reversal, Neurodevelopmental Delay, Epilepsy and Optic Atrophy](#). *Genes (Basel).* 2024;15(7):901. doi:10.3390/genes15070901
3. Fang X, Ma M, Rong W, et al. [Exome sequencing confirms the clinical diagnosis of both joubert syndrome and klinefelter syndrome with keratoconus in a han Chinese family](#). *Front Genet.* 2024;15:1417584. doi:10.3389/fgene.2024.1417584
4. Zeng Y, Ding H, Wang X, et al. [High positive predictive value of CNVs detected by clinical exome sequencing in suspected genetic diseases](#). *J Transl Med.* 2024;22(1):644. doi:10.1186/s12967-024-05468-1
5. Harrison PW, Amode MR, Austine-Orimoloye O, et al. [Ensembl 2024](#). *Nucleic Acids Res.* 2024;52(D1):D891-D899. doi:10.1093/nar/gkad1049
6. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. [Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model](#). *BMC Med Genomics.* 2018;11(1):91. doi:10.1186/s12920-018-0404-4
7. Greenwood Genetic Center. [GGCC reduces turn around time on genomic analysis by 75% with Emedgene's AI platform](#). [ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform](https://ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform). Publicado em 12 de setembro de 2019. Acessado em 14 de agosto de 2024.
8. Dados em arquivo. Illumina, Inc. 2024.



+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-01453 PTB v1.0