

Design de painel personalizado para o Illumina Complete Long Read Prep with Enrichment, Human

Enriquecimento altamente flexível e direcionado de leitura longa para genomas humanos



Introdução

Ao realizar sequenciamento do genoma completo (WGS) humano, várias regiões podem ser difíceis de mapear com leituras curtas isoladamente. O sequenciamento de leitura longa pode complementar os dados WGS de leitura curta padrão para ajudar a abordar essas regiões desafiadoras. A tecnologia Illumina Complete Long Reads usa um fluxo de trabalho padrão de sequenciamento de última geração (NGS) para gerar sequências de leitura longa contíguas em sistemas de sequenciamento Illumina com um único pipeline de análise (Figura 1).¹⁻³ O Illumina Complete Long Read Prep with Enrichment, Human adiciona uma abordagem direcionada para sequenciamento de leitura longa mais econômico.* A química de enriquecimento do Illumina Complete Long Read oferece alta flexibilidade com alvos e design de sonda para ajudar a resolver regiões difíceis de mapear ou fornecer informações adicionais do sequenciamento em fases.

Projetar painéis de sonda de enriquecimento para leituras longas

Illumina Complete Long Read Prep with Enrichment, Human usa uma estratégia de projeto de sonda diferente para capturar fragmentos mais longos (~7–10 kb) do que a abordagem normalmente usada para capturar fragmentos curtos (~200–500 bp). O software Illumina DesignStudio™ é uma ferramenta gratuita e fácil de usar para projetar painéis de sondas de enriquecimento.

* Requer ≥ 30× dados padrão de leitura curta de WGS da mesma amostra para análise. Os arquivos FASTQ da amostra de uma corrida anterior podem ser usados.

O algoritmo DesignStudio considera o conteúdo do GC, a especificidade do alvo e o espaçamento da sonda, ou seja, quantas sondas estarão na região alvo. O espaçamento padrão para painéis de enriquecimento de leitura curta de 120 mer é uma janela de sonda de 250–350 bp. Para o design do painel de enriquecimento de leitura longa, o espaçamento da sonda foi testado em vários comprimentos e uma janela de um quilobase foi considerada ideal para captura econômica e de alta eficiência.

A eficácia do enriquecimento da hibridização depende muito da especificidade da sonda. A porcentagem de enriquecimento no alvo afeta diretamente a quantidade de sequenciamento necessária para atingir a profundidade de cobertura do alvo. É mais difícil obter alta especificidade para regiões repetitivas. No entanto, o uso de uma janela de sonda maior permite maior flexibilidade para excluir sondas de baixo desempenho, evitar regiões de repetição (até o tamanho da janela de 1 kb) e manter a eficiência de enriquecimento com menos sondas (Figura 2). O algoritmo do DesignStudio pode usar essas considerações para recomendar a colocação da sonda. Painéis de terceiros devem usar diretrizes semelhantes para obter o melhor desempenho e eficiência de custos. O espaçamento padrão da sonda de enriquecimento também é totalmente compatível.

Flexibilidade no projeto da sonda e na estratégia de destino

Illumina Complete Long Read Prep with Enrichment, Human oferece alta flexibilidade para escolher e projetar painéis de sonda personalizados para atender às metas do estudo. As regiões-alvo individuais podem abranger bases únicas de até centenas de quilobases. O tamanho total do painel pode variar de painéis personalizados tão pequenos quanto 2,5 Mb até > 95 Mb.



Figura 1: Parte de um fluxo de trabalho integrado: acesse dados WGS de leitura longa direcionados e econômicos usando uma preparação de biblioteca escalável e otimizada com protocolo de enriquecimento, química comprovada de sequenciamento Illumina e análise secundária DRAGEN. Requer ≥ 30× dados padrão de leitura curta de WGS da mesma amostra para análise. Os arquivos FASTQ da amostra de uma corrida anterior podem ser usados.

Os pesquisadores podem usar leituras longas direcionadas para melhorar a cobertura em regiões específicas conhecidas por terem baixa capacidade de mapeamento com dados de leitura curta. Alternativamente, leituras longas podem ser direcionadas para cobrir genes inteiros até regiões multigene longas para permitir o faseamento de variantes e a identificação de haplótipos.

Vários painéis pré-projetados estão disponíveis na ferramenta DesignStudio (Tabela 1). Esses painéis têm como alvo genes desafiadores clinicamente relevantes (CMRG),⁴ genes comumente direcionados por ensaios de testes farmacogenéticos (PGx),^{5 a 7} genes na lista de achados secundários do American College of Medical Genetics and Genomics (ACMG) (ACMG SF v3.1),⁸ ou a região completa do complexo principal de histocompatibilidade (MHC).⁹ O Painel Abrangente Humano da Illumina, que visa principalmente regiões discretas de baixa cobertura dentro de genes codificadores de proteínas, também está disponível como um painel pré-design ou pronto para envio pré-fabricado (Illumina, nº de catálogo 20113836).^{10,11} O software DesignStudio é compatível com o projeto de painéis personalizados a partir de arquivos BED† ou com a modificação de pré-projetos existentes.

Profundidade de sequenciamento recomendada para painéis de sonda personalizados

Illumina Complete Long Read Prep with Enrichment, Human oferece desempenho altamente consistente e robusto. Para os painéis pré-projetados testados, o desempenho ideal foi alcançado com ~1,5 Gb de dados de sequência (~5 milhões de leituras do tipo paired-end) por 1 Mb de tamanho do painel alvo (Figura 3). Para painéis recém-projetados com desempenho desconhecido, 3 Gb de dados de sequência (~10 milhões de leituras tipo paired-end) por 1 Mb de tamanho de painel alvo é um ponto de partida recomendado, sujeito a redução com otimização adicional.

Cobertura de alta precisão e faseamento de regiões difíceis

Painéis de sondas de enriquecimento de leitura longa focados em melhorar regiões específicas de baixa cobertura, como o Painel Humano Abrangente Illumina e o painel CMRG, aumentam a precisão da identificação de variantes em regiões desafiadoras direcionadas (Figura 4). O enriquecimento de leitura longa com o painel CMRG também ajuda a melhorar a integridade da cobertura e detecção de variantes em regiões codificantes de proteínas (Figura 5, Figura 6).

† BED, formato de dados extensível do navegador.

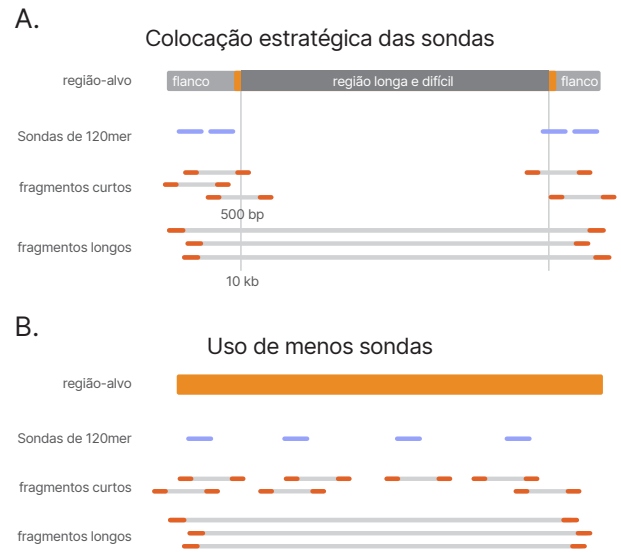


Figura 2: Hibridização de fragmentos longos aumenta a eficiência do enriquecimento: A hibridização de fragmentos longos oferece vantagens em relação à captura de fragmentos curtos, incluindo (A) a colocação estratégica de sondas fora de regiões difíceis para o projeto de sondas, como conteúdo extremo de GC, baixa complexidade ou repetições, e (B) menos sondas são necessárias para capturar cada região alvo. O algoritmo DesignStudio pesquisa em regiões-alvo em seções de 1 kb quanto ao conteúdo ideal de GC e regiões de maior especificidade para colocar sondas.

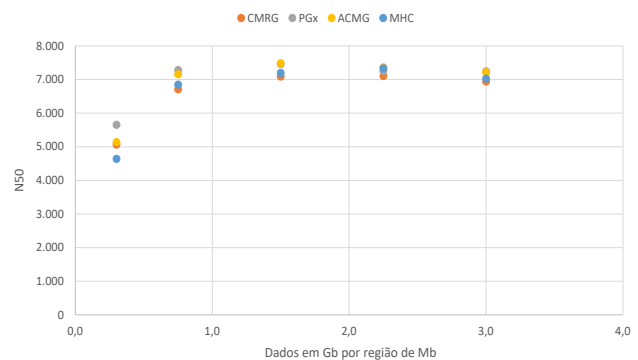


Figura 3: Requisitos de sequenciamento para painéis de sonda personalizados: a titulação dos dados de sequenciamento necessários para o N50 máximo mostra que 1,5 Gb (~5M de leitura tipo paired-end) por região alvo Mb resolve com eficiência as regiões-alvo para a geração de dados do Illumina Complete Long Read.

Tabela 1: Painéis de sonda de enriquecimento pré-projetados para Illumina Complete Long Read Prep with Enrichment, Human

Painéis ^a	Painel CMRG	Painel PGx	Painel ACMG	Painel MHC
Genes direcionados	391 genes clinicamente relevantes conhecidos por serem difíceis de resolver com leituras curtas ⁵	98 genes comumente direcionados por ensaios de testes farmacogenéticos ^{6 a 8}	78 genes exclusivos da lista de achados secundários do ACMG (ACMG SF v3.1) ⁹	> 140 genes na região completa do MHC no conjunto GRCh38.p14 ¹⁰
Tamanho da região alvo ^b	22,5 Mb	8,1 Mb	7 Mb	4,9 Mb
Saída de sequenciamento por amostra ^c	~67,5 Gb	~24,3 Gb	~21 Gb	~14,7 Gb
Nº de sondas	~22,5 mil	~8,2 mil	~6,9 mil	~5,0 mil
N50 ^d	6,1 kb	7,3 kb	7,3 kb	7,3 kb
Bloco de fase N50 ^{d,e}	82,8 kb	94,4 kb	94,4 kb	357 kb
Tamanho médio da região alvo ^e	58 kb	83 kb	88 kb	5.000 kb
Uniformidade ^{d,f}	97,9%	99,0%	99,5%	97,8%
Enriquecimento de leitura "padded" (PRE) ^{d,f}	80,1%	79,3%	66,3%	67,5%
% de SNV heterozigótica em fases ^d	98,9%	98,9%	99,6%	98,6%

- a. CMRG, genes clinicamente relevantes desafiadores; PGx, farmacogenômica; ACMG, American College of Medical Genetics and Genomics; MHC, complexo principal de histocompatibilidade.
- b. O tamanho da região-alvo é a soma dos comprimentos de localização da sonda "padded", mesclados onde se sobrepõem.
- c. Requer uma corrida de sequenciamento de 2 x 150 bp e leituras tipo paired-end de 5MA10M (~1,5A3 Gb de dados) por região alvo de Mb, gerando aproximadamente 30x a cobertura final do Illumina Complete Long Reads. Os requisitos de dados personalizados do painel por amostra são apenas um ponto de partida recomendado. Os usuários podem otimizar os dados alocados com base no desempenho do painel.
- d. Dados gerados usando 50 ng de DNA genômico de HG002 (Coriell, nº de catálogo NA24385). O desempenho pode variar com a entrada de DNA e a qualidade da amostra.
- e. Os tamanhos dos blocos de fase são limitados aos tamanhos de regiões-alvo contíguas individuais.
- f. Uniformidade da cobertura calculada como % > 0,2 * média. PRE calculado como 100 * (leituras alinhadas de alvo "padded"/leituras alinhadas totais).

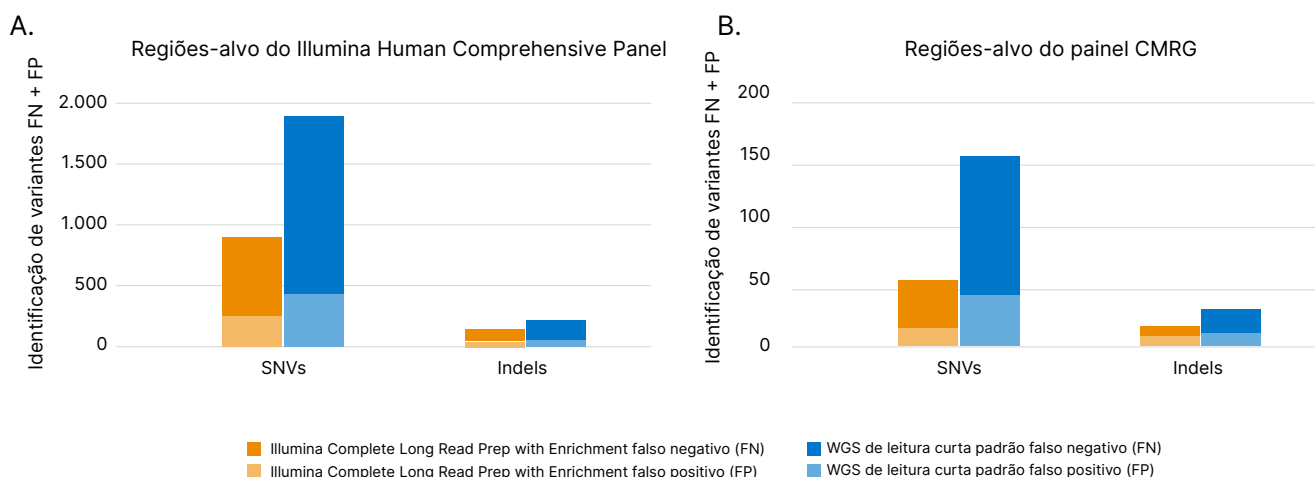


Figura 4: Leituras longas direcionadas para melhorar a precisão da identificação de variantes em regiões desafiadoras: a variante falso negativa (FN) mais falso positiva (FP) exige variantes de único nucleotídeo (SNVs) e indels em regiões gênicas HG002 direcionadas pelo (A) Painel Abrangente Humano da Illumina ou (B) painel CMRG, usando o Illumina Complete Long Read Prep with Enrichment (laranja) em comparação com WGS de leitura curta padrão (azul).



Figura 5: Leituras longas direcionadas melhoram as regiões de baixa cobertura: gráficos do Integrated Genomics Viewer (IGV) do sequenciamento de leitura longa do *HBG1* usando Illumina Complete Long Read Prep, Human WGS (superior) e Illumina Complete Long Read Prep com Enrichment, Human e o painel CMRG (intermediário) em comparação com WGS de leitura curta padrão (parte inferior).

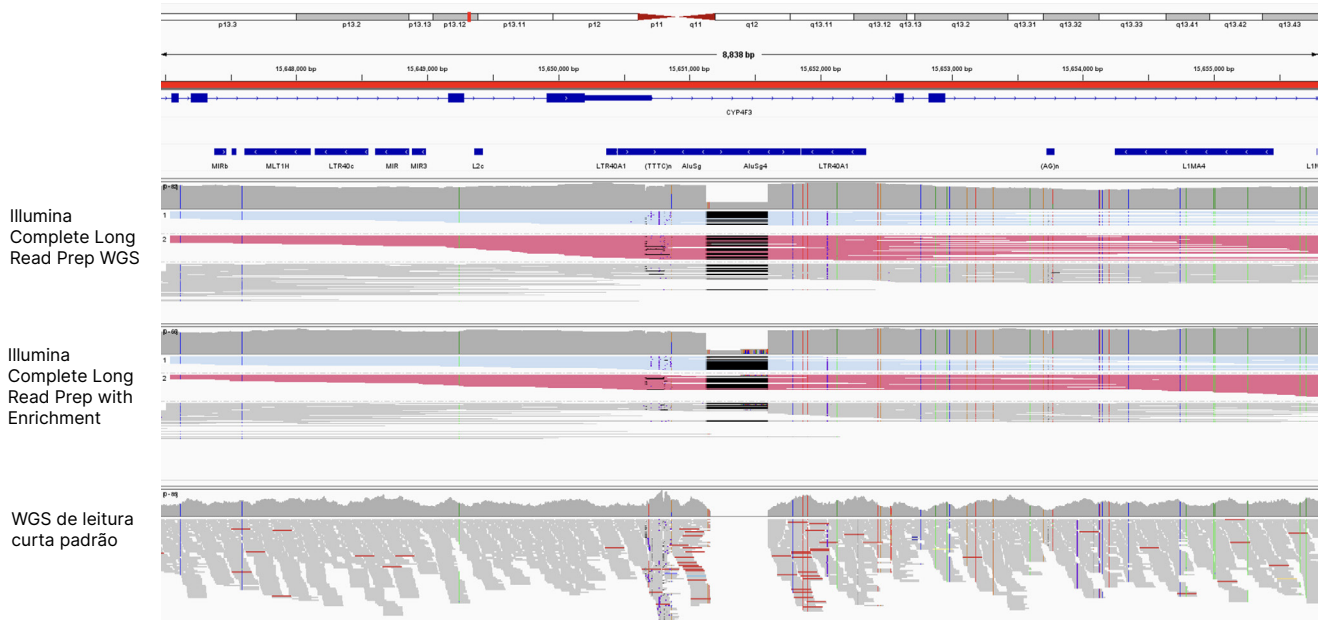
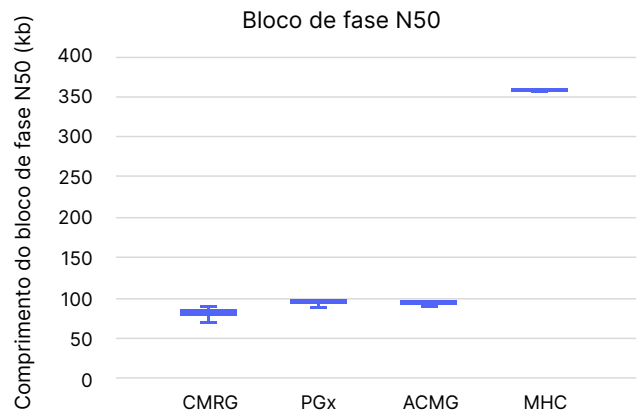


Figura 6: Resolução clara de bordas de exclusão com leituras longas direcionadas: gráficos IGV do sequenciamento de leitura longa e faseamento do *CYP4F3* usando Illumina Complete Long Read Prep, Human WGS (superior) e Illumina Complete Long Read Prep com Enrichment, Human e o painel CMRG (intermediário) em comparação com WGS de leitura curta padrão (parte inferior). Alelo 1 em azul, alelo 2 em rosa.

Blocos de fase longa para resolver haplótipos

O bloco de fase N50[‡] de cada painel está relacionado ao comprimento contíguo das regiões alvo (Figura 7, Tabela 1). Os painéis CMRG, PGx e ACMG são projetados para atingir genes de comprimento total de interesse e produziram um bloco de fase N50 médio de ~80–95 kb para permitir a fase completa de alelos heterozigóticos (Figura 8). O painel de MHC tem como alvo uma única região contígua de ~4,9 Mb e produziu um bloco de fase médio N50 de mais de 350 kb para permitir a resolução da região gênica de comprimento total (Figura 9).



‡ O bloco de fase N50 reflete o comprimento do bloco mais curto da sequência contígua a 50% do comprimento total do conjunto das regiões-alvo.

Figura 7: O bloco de fase N50 depende do comprimento das regiões-alvo contíguas: os painéis CMRG, PGx e ACMG têm como alvo genes de comprimento total de interesse e produziram o bloco de fase N50 médio de ~80 a 95 kb. O painel do MHC tem como alvo a região gênica completa do complexo principal de histocompatibilidade e produziu um bloco de fase N50 médio de mais de 350 kb. O tamanho médio da região alvo é de 58 kb para o painel CMRG, 83 kb para o painel PGx, 88 kb para o painel ACMG e 5.000 kb para o painel MHC.

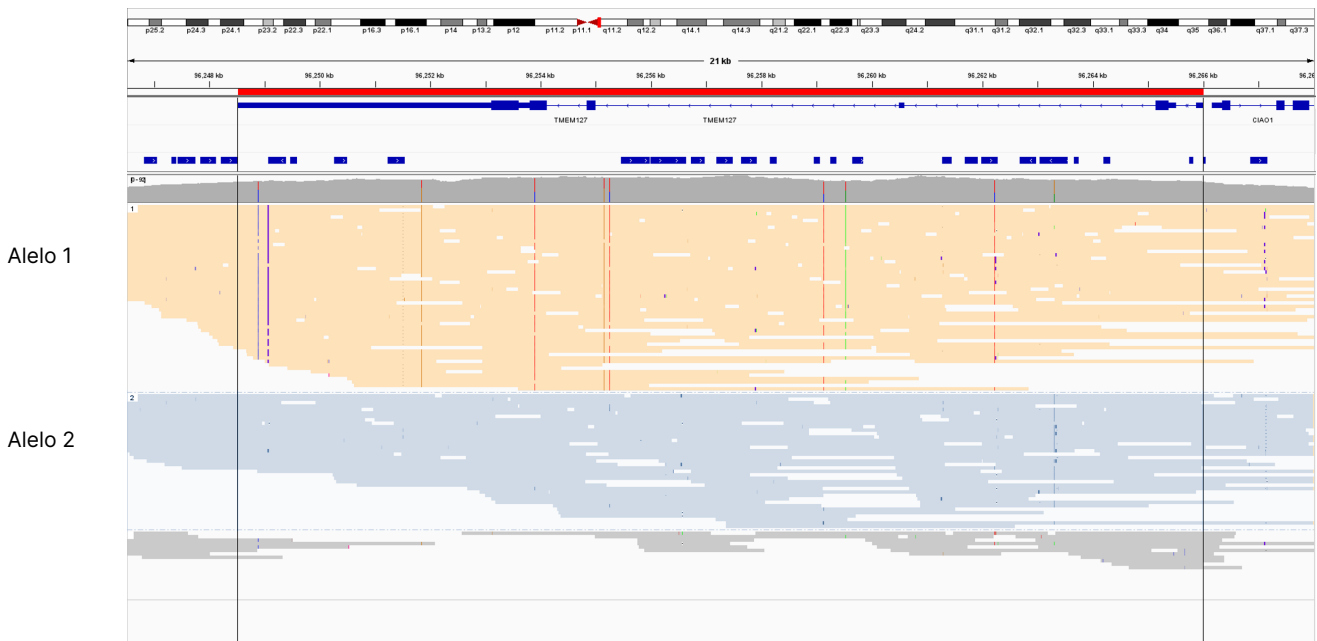


Figura 8: Leituras longas direcionadas permitem o faseamento de regiões com SNVs heterozigóticas: gráficos do IGV do sequenciamento de leitura longa mostram faseamento completo em um bloco de fase para *TMEM127*, um gene de 21 kb, usando Illumina Complete Long Read Prep com Enrichment, Human e o painel ACMG. Alelo 1 em amarelo. Alelo 2 em azul.

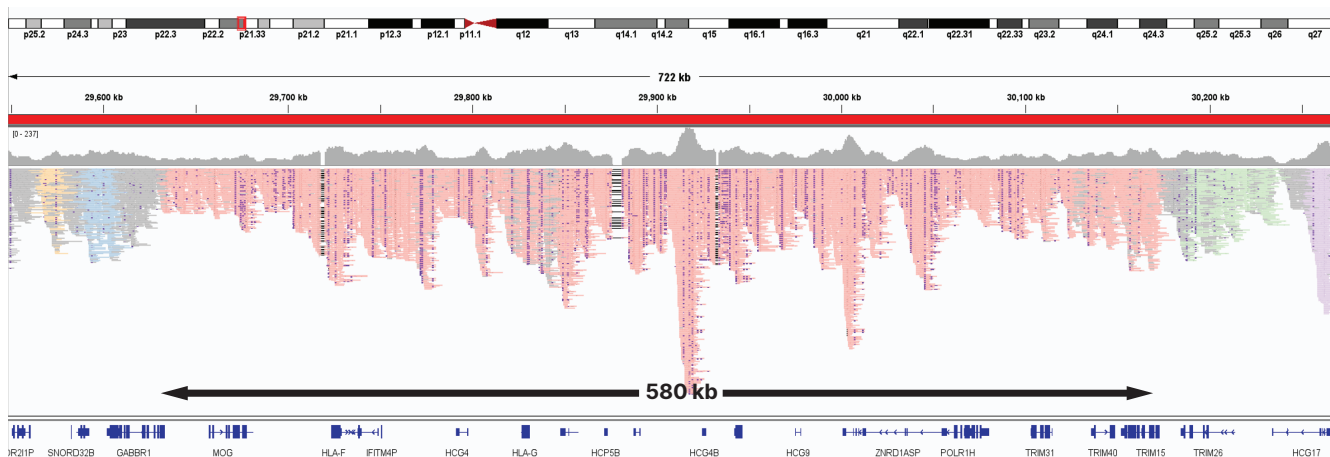


Figura 9: Leituras longas direcionadas ajudam a resolver haplótipos em genes polimórficos: gráficos do IGV de sequenciamento de leitura longa usando Illumina Complete Long Read Prep with Enrichment, Human. Faseamento sobre uma região de 722 kb no loco de MHC. Uma região de 580 kb (rosa) é encapsulada em um bloco de fase.

Resumo

O Illumina Complete Long Read Prep with Enrichment, Human complementa o comprovado WGS de leitura curta do Illumina e concentra o sequenciamento de leitura longa onde ele fornece o maior valor. Os pesquisadores têm alta flexibilidade para escolher painéis pré-projetados ou usar o algoritmo DesignStudio para projetar painéis personalizados para enriquecimento direcionado de leitura longa. Os painéis de sondas de enriquecimento podem ser direcionados para melhorar a cobertura ou adicionar insights, ao fasear genes inteiros para obter WGS econômicos e altamente precisos com uma solução de fluxo de trabalho completa.

Saiba mais

[Illumina Complete Long Read Prep with Enrichment, Human](#)

[Ferramenta de projeto de ensaio DesignStudio](#)

[Tecnologia de sequenciamento de leitura longa](#)

Referências

1. Illumina. Illumina Complete Long Read Prep, Human data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf). Publicado em 2022. Acessado em 22 de setembro de 2023.
2. Illumina. Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human technical note. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf). Publicado em 2022. Acessado em 22 de setembro de 2023.
3. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. <https://www.illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html>. Publicado em 2023. Acessado em 22 de setembro de 2023.
4. Wagner J, Olson ND, Harris L, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.* 2022;40(5):672-680. doi:10.1038/s41587-021-01158-1
5. PharmGKB. VIPs: Very Important Pharmacogenes. [pharmgkb.org/vips](https://www.pharmgkb.org/vips). Acessado em 22 de setembro de 2023.
6. National Library of Medicine. GTR: Genetic Testing Registry. Precision HealthPGx Panel (25 Genes). [ncbi.nlm.nih.gov/gtr/tests/593428/](https://www.ncbi.nlm.nih.gov/gtr/tests/593428/). Atualizado em 29 de novembro de 2022. Acessado em 22 de setembro de 2023.

7. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn.* 2016;18(1):109-123. doi:10.1016/j.jmoldx.2015.08.005
8. Miller DT, Lee K, Abul-Husn NS, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2022;24(7):1407-1414. doi:10.1016/j.gim.2022.04.006
9. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var.* 2022;9(1):49. doi:10.1038/s41439-022-00226-5
10. Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. Publicado em 2021. Acessado em 30 de agosto de 2023.
11. Illumina. Illumina Human Comprehensive Panel data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf). Publicado em 2024. Acessado em 26 de janeiro de 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.
M-GL-02189 PTB v1.0