

# NextSeq™ 550Dx Instrument

PARA USO EM DIAGNÓSTICO IN VITRO  
SOMENTE PARA EXPORTAÇÃO

N.º do catálogo 20005715

## Uso previsto

O NextSeq 550Dx Instrument é destinado ao sequenciamento de bibliotecas de DNA quando usado em ensaios para diagnóstico *in vitro*. O NextSeq 550Dx Instrument deve ser usado com reagentes específicos para diagnóstico *in vitro* registrados, certificados ou aprovados e com software analítico.

## Princípios do procedimento

O Illumina NextSeq 550Dx Instrument se destina a fazer o sequenciamento de bibliotecas de DNA com ensaios de diagnóstico *in vitro* e deve ser usado por equipes treinadas no uso de procedimentos de diagnóstico *in vitro* conduzidos em um laboratório clínico. Para a entrada, o NextSeq 550Dx usa bibliotecas geradas de DNA em que índices de amostra e sequências de captura são adicionados a alvos amplificados. As bibliotecas de amostras são capturadas em uma lâmina de fluxo e sequenciadas no instrumento com o uso da química de sequenciamento por síntese (SBS, sequencing by synthesis). A química de SBS usa um método de terminador reversível para detectar bases de nucleotídeo únicas identificadas por fluorescência, à medida que são incorporadas aos filamentos crescentes de DNA. O software Real-Time Analysis (RTA) faz a análise da imagem e a identificação de base e atribui uma pontuação de qualidade a cada base de cada ciclo de sequenciamento. Quando a análise primária é concluída, a análise secundária pode ser executada no instrumento para processar as identificações de bases. O NextSeq 550Dx usa módulos diferentes de análise secundária, dependendo do fluxo de trabalho. Para os módulos de variante somática ou de linha genética, o processamento inclui demultiplexação, geração de arquivos FASTQ, alinhamento, identificação de variantes e geração de arquivos em formato de identificação de variantes (VCF e gVCF). Os arquivos VCF e gVCF contêm informações sobre variantes encontradas em posições específicas em um genoma de referência.

## Configuração de inicialização dupla

O NextSeq 550Dx inclui uma configuração de inicialização dupla para permitir o uso do instrumento em modo diagnóstico (Dx) ou no modo somente para uso em pesquisa (RUO, Research Use Only). Ensaios de sequenciamento para diagnóstico *In vitro*, incluindo os módulos de variante de linha genética e somática, são executados em modo diagnóstico. Só podem ser usados reagentes de sequenciamento para diagnóstico *in vitro* (IVD, In Vitro Diagnosis) no modo diagnóstico. As características de desempenho e as limitações de procedimento do NextSeq 550Dx Instrument foram estabelecidas por meio dos módulos de variante de linha genética e somática no modo diagnóstico.

## Limitações do procedimento

- 1 Para uso em diagnóstico *in vitro*.
- 2 Os módulos de variantes somáticas e de linha genética, quando usados com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) ou com o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), têm a capacidade de produzir:
  - ▶ Saída de sequenciamento  $\geq 90$  gigabases (Gb)
  - ▶ Duração da leitura (em execução tipo paired-end) 2 x 150 pares de base (bp)

- ▶ Bases maiores ou iguais a Q30  $\geq 75\%$  com duração da leitura de 2 x 150 bp  
Setenta e cinco por cento das bases ou mais têm pontuações de qualidade da escala Phred  $\geq 30$ , indicando identificação de bases com exatidão acima de 99,9%.
- 3 Leituras com indels (inserções, deleções ou combinações) em que a duração do conteúdo é  $>25$  pares de bases (bp) não são alinhadas pelo software do ensaio. Consequentemente, indels com duração  $>25$  bp não são detectáveis pelo software do ensaio.
- 4 O software do ensaio pode não alinhar leituras de amplicons com um conteúdo de variantes extremo, o que resulta na região ser relatada como do tipo selvagem. Esse conteúdo extremo abrange:
  - ▶ Leituras contendo mais de três indels.
  - ▶ Leituras com comprimento mínimo de 30 bp com conteúdo de variante de único nucleotídeo (SNV)  $>4\%$  do comprimento total alvo de amplicon (excluindo regiões de sonda).
  - ▶ Leituras com comprimento  $<30$  bp com conteúdo de SNV  $>10\%$  do comprimento total do amplicon (incluindo regiões de sonda).
- 5 Grandes variantes, incluindo variantes de vários nucleotídeos (MNVs, multi-nucleotide variants) e grandes indels, podem ser reportadas como variantes menores separadas no arquivo VCF de saída.
- 6 As variantes de deleção poderão ser filtradas ou poderão se perder durante a extensão de dois amplicons em bloco se o comprimento da deleção for maior ou igual à sobreposição entre os amplicons em bloco.
- 7 O sistema não pode detectar indels se estes ocorrerem diretamente adjacentes a um primer e não houver qualquer amplicon sobreposto. Para regiões com amplicons sobrepostos, o ensaio não consegue detectar deleções quando a região de sobreposição é menor que o tamanho da deleção a ser detectada. Por exemplo, se a região de sobreposição entre dois amplicons adjacentes for duas bases, o ensaio não conseguirá detectar qualquer deleção que inclua essas duas bases. Pode ser detectada uma deleção de base única em cada uma dessas bases.
- 8 Da mesma forma que em qualquer fluxo de trabalho de preparação de bibliotecas com base em hibridização, os polimorfismos, mutações, inserções ou deleções subjacentes nas regiões de ligação de oligonucleotídeos podem afetar os alelos que estão sendo sondados e as identificações feitas durante o sequenciamento. Por exemplo:
  - ▶ Uma variante em fase com uma variante na região do primer pode não ser amplificada, resultando em um falso negativo.
  - ▶ Variantes na região do primer podem evitar a amplificação do alelo de referência, resultando em uma identificação de variante homocigótica incorreta.
  - ▶ Variantes de indel na região do primer podem causar uma identificação falso positiva no final da leitura adjacente ao primer.
- 9 Indels podem ser filtrados devido à inclinação de filamento se ocorrerem próximo ao final de uma leitura e forem submetidos a soft-clipping durante o alinhamento.
- 10 MNVs pequenos não foram validados e são somente reportados no Módulo de variante somática.
- 11 As deleções são reportadas no arquivo VCF na coordenada da base precedente segundo o formato VCF. Portanto, leve em consideração as variantes adjacentes antes de relatar que uma determinada identificação de bases é uma referência homocigótica.
- 12 Limitações específicas de linha genética:
  - ▶ O NextSeq 550Dx Instrument, com o Módulo de variante de linha genética do Local Run Manager para NextSeq 550Dx, destina-se a apresentar resultados qualitativos para identificação de variante de linha genética (por exemplo, homocigótica, heterocigótica, tipo selvagem).
  - ▶ Quando usado com o Módulo de variante de linha genética, a cobertura mínima por amplicon necessária para identificação precisa de variante é de 150x. Como resultado, 150 fragmentos de DNA de sustentação são necessários, o que é equivalente a 300 leituras tipo paired-end sobrepostas. O número de amostras e o número total de bases visadas afetam a cobertura. O conteúdo GC e outro conteúdo genômico podem afetar a cobertura.
  - ▶ A variação do número de cópias pode afetar o fato de uma variante ser identificada como homocigótica ou heterocigótica.

- ▶ As variantes de determinado contexto repetitivo são filtradas nos arquivos VCF. O filtro de repetição RMxN é usado para filtrar variantes se toda ou parte da sequência de variante estiver presente repetidamente no genoma de referência adjacente à posição de variante. Para identificação de variante de linha genética, pelo menos nove repetições na referência são necessárias para que uma variante seja filtrada. Somente repetições com comprimento de até 5 bp são consideradas (R5x9).
  - ▶ Um indel e um SNV em um único locus podem resultar em apenas uma variante ser relatada.
- 13 Limitações específicas somáticas.
- ▶ O NextSeq 550Dx Instrument, com o Local Run Manager Somatic Variant Module for NextSeq 550Dx, se destina a apresentar resultados qualitativos para identificação de variante somática (por exemplo, presença de uma variante somática com frequência de variante maior ou igual a 0,026 com limite de detecção de 0,05).
  - ▶ Quando usado com o Módulo de variante somática, a cobertura mínima por amplicon necessária para identificação de variante precisa é 450x por pool de oligonucleotídeos. Como resultado, 450 fragmentos de DNA de sustentação são necessários por pool de oligonucleotídeos, o que é equivalente a 900 leituras do tipo paired-end sobrepostas. O número de amostras e o número total de bases visadas afetam a cobertura. O conteúdo GC e outro conteúdo genômico podem afetar a cobertura.
  - ▶ Para identificação de variante somática, pelo menos seis repetições na referência são necessárias para que a variante seja filtrada, e apenas repetições com comprimento de até 3 bp são consideradas (R3x6).
  - ▶ O módulo de variantes de linha somática não consegue diferenciar entre variantes de linha genética e variantes de linha somática. O módulo foi projetado para detectar variantes em um intervalo de frequências de variantes, mas a frequência de variantes não pode ser usada para diferenciar variantes de linha somática de variantes de linha genética.
  - ▶ O tecido normal no espécime afeta a detecção de variantes. O limite reportado de detecção é baseado em uma frequência de variante em relação ao DNA total extraído tanto de tecido tumoral quanto de tecido normal.

## Componentes do produto

- 1 NextSeq 550Dx Instrument (n.º do catálogo 20005715)
- 2 Componentes do software para o NextSeq 550Dx Instrument, incluindo o seguinte:

Aplicativo do software	Função	Descrição
NextSeq 550Dx Operating Software (NOS)	Controla a operação do instrumento	O aplicativo de software NOS gerencia a operação do instrumento durante o sequenciamento e gera imagens que serão usadas pelo software Real-Time Analysis (RTA).
Real-time Analysis Software (RTA)	Executa a análise primária	O aplicativo de software RTA converte as imagens geradas pelo NOS para cada bloco por ciclo de execução do sequenciamento em arquivos de identificação de bases, que são entradas para os módulos de análise do Local Run Manager. O aplicativo de software RTA não contém uma interface de usuário.
Local Run Manager	Interface para seleção de módulos	O software Local Run Manager é uma solução integrada do instrumento para gerenciamento pelo usuário, seleção do módulo de análise apropriado e monitoramento do status.
Módulo de variante somática	Executa a análise secundária	Esse software do módulo de análise do Local Run Manager processa identificações de bases por meio de análise secundária. O processamento inclui demultiplexação, geração de arquivos FASTQ, alinhamento, identificação de variantes e geração de relatórios. O identificador de variantes (Pisces) gera arquivos VCF que contêm informações sobre as variantes encontradas em posições específicas de um genoma de referência e inclui a frequência da variante medida.

Aplicativo do software	Função	Descrição
Módulo de variante de linha genética	Executa a análise secundária	Esse software do módulo de análise do Local Run Manager processa identificações de bases por meio de análise secundária. O processamento inclui demultiplexação, geração de arquivos FASTQ, alinhamento, identificação de variantes e geração de relatórios. O identificador de variantes (Pisces) gera arquivos VCF que contêm informações sobre as variantes encontradas em posições específicas de um genoma de referência e identifica cada variante como heterozigótica ou homozigótica.

## Condições de operação

Elemento	Especificação
Temperatura	Mantenha a temperatura do laboratório entre 19 °C e 25 °C (22 °C $\pm$ 3 °C). Esta é a temperatura de operação do instrumento. Durante uma execução, não permita que a temperatura ambiente varie mais do que $\pm$ 2 °C.
Umidade	Mantenha uma umidade relativa não condensada de 20% a 80%.

## Equipamento e materiais

### Equipamento e materiais necessários, vendidos separadamente

NextSeq 550Dx High Output Reagent Kit v2.5 (75 cycles), n.º do catálogo 20028870

NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), n.º do catálogo 20028871

### Equipamento e materiais necessários, não fornecidos

#### Materiais de consumo fornecidos pelo usuário para execuções do sequenciamento

Material de consumo	Fornecedor	Finalidade
Compressa com álcool isopropílico 70% ou etílico 70%	VWR, n.º do catálogo 95041-714 (ou equivalente) Fornecedor de itens de uso comum do laboratório	Limpeza da lâmina de fluxo e uso geral
Lenço para laboratório com poucos fiapos	VWR, n.º do catálogo 21905-026 (ou equivalente)	Limpeza da lâmina de fluxo e uso geral

#### Materiais de consumo fornecidos pelo usuário para manutenção do instrumento

Material de consumo	Fornecedor	Finalidade
NaOCl, 5% (hipoclorito de sódio)	Sigma-Aldrich, n.º do catálogo 239305 (ou equivalente aprovado para uso em laboratório)	Limpar o instrumento usando o material de limpeza manual após a execução; diluído a 0,12%.
Tween 20	Sigma-Aldrich, catálogo n.º P7949	Limpar o instrumento usando opções de limpeza manual; diluído a 0,05%.
Água, aprovada para uso em laboratório	Fornecedor de itens de uso comum do laboratório	Limpar o instrumento (limpeza manual).
Filtro de ar	Illumina, n.º do catálogo 20022240	Limpar o ar que o instrumento recebe para o resfriamento.

## Orientações para água aprovada para uso em laboratório

Utilize sempre água aprovada para uso em laboratório ou água deionizada para realizar procedimentos com instrumentos. Nunca use água da torneira. Utilize apenas os seguintes tipos de água ou equivalentes:

- ▶ Água deionizada
- ▶ Illumina PW1
- ▶ Água de 18 Megohm (M $\Omega$ )
- ▶ Água Milli-Q
- ▶ Água Super-Q
- ▶ Água para biologia molecular

## Advertências e precauções

**CUIDADO** A lei federal restringe este dispositivo a ser vendido por ou mediante a ordem de um médico ou outro profissional da área médica licenciado pela lei do estado no qual atua, para usar ou designar o uso do dispositivo.

- 1 **Alguns componentes dos reagentes fornecidos pela Illumina para uso com o NextSeq 550Dx Instrument contêm substâncias químicas potencialmente perigosas. Podem ocorrer ferimentos por meio de inalação, ingestão e contato com a pele ou com os olhos. Use equipamento de proteção, incluindo proteção para os olhos, luvas e jaleco, apropriado para risco de exposição. Manuseie os reagentes usados como resíduo químico e descarte-os de acordo com as leis e regulamentações regionais, nacionais e locais aplicáveis.** Para obter informações ambientais, de saúde e segurança adicionais, consulte as fichas de dados de segurança (SDS) em [support.illumina.com/sds.html](https://support.illumina.com/sds.html).
- 2 Imediatamente informe à Illumina e às autoridades competentes dos estados membros em que o usuário e o paciente estiverem estabelecidos sobre quaisquer incidentes graves relacionados a este produto.
- 3 Manuseie todas as amostras de sangue como se fossem infecciosas para o vírus da imunodeficiência humana (HIV), o vírus da hepatite B humana (HBV) e outros agentes patogênicos transmitidos pelo sangue (precauções universais).
- 4 Se os procedimentos não forem seguidos conforme o estabelecido, poderá haver resultados com erro ou uma redução significativa na qualidade da amostra.
- 5 Adote precauções de rotina no laboratório. Não utilize a pipeta com a boca. Não coma, beba ou fume em áreas de trabalho designadas. Use luvas descartáveis e jalecos para laboratório ao manusear espécimes e kits de reagentes. Lave bem as mãos depois de manusear espécimes e kits de reagentes.
- 6 Práticas adequadas e boa higiene no laboratório são necessárias para evitar que os produtos de PCR contaminem reagentes, instrumentos e amostras de DNA genômico. A contaminação por PCR pode causar resultados inexatos e não confiáveis.
- 7 Para evitar a contaminação, certifique-se de que as áreas de pré-amplificação e pós-amplificação possuam equipamentos e materiais de consumo exclusivos (p. ex., pipetas, pontas de pipeta, blocos de calor, agitadores e centrífuga).
- 8 O índice para o emparelhamento de amostras deve corresponder exatamente ao layout da placa impressa. O Local Run Manager preenche automaticamente os primers de índice associados aos nomes das amostras, quando inserido no módulo. O usuário é aconselhado a verificar os primers de índice associados às amostras antes de iniciar a execução do sequenciamento. A incompatibilidade entre a planilha de amostras e o layout da placa provoca a perda da identificação positiva da amostra e a geração de relatórios com resultados incorretos.
- 9 A instalação de software antivírus fornecido pelo usuário é altamente recomendável para proteger o computador contra vírus. Consulte as instruções de instalação no manual do usuário.
- 10 Não opere o NextSeq 550Dx se um dos painéis tiver sido removido. Operar o instrumento sem um dos painéis pode causar exposição à tensão de linha e a tensões CC.

- 11 Não toque no estágio da lâmina de fluxo no compartimento da lâmina de fluxo. O aquecedor desse compartimento opera entre 22 °C e 95 °C e pode causar queimaduras.
- 12 O instrumento pesa aproximadamente 84 kg e pode causar sérios ferimentos se cair ou se for manuseado de forma imprópria.

## Instruções de uso

As instruções de uso a seguir abrangem a execução dos módulos de variante de linha genética e somática em modo diagnóstico no NextSeq 550Dx Instrument usando o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) ou o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

## Inserir as informações da execução

Para obter instruções detalhadas, consulte o NextSeq 550Dx Instrument Reference Guide (Guia de referência do NextSeq 550Dx Instrument) (documento n.º 1000000009513) e o guia do módulo do Local Run Manager aplicável.

### Definir parâmetros

- 1 Faça log-in no Local Run Manager.
- 2 Selecione **Create Run** (Criar execução) e depois **Somatic Variant** (Variante somática) ou **Germline Variant** (Variante de linha genética).
- 3 Informe um nome que identifique a execução desde o sequenciamento até a análise.  
Use caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 4 **[Opcional]** Informe uma descrição para ajudar a identificar a execução.  
Use caracteres alfanuméricos, espaços, sublinhados ou travessões.
- 5 Selecione o número de amostras e um conjunto de índices na lista suspensa.  
Considere as seguintes informações quando fizer uma seleção.
  - ▶ A lista suspensa contém números de amostras com um conjunto de índices. Por exemplo, "24-Set 1" indica 24 amostras a serem testadas, com índices do conjunto de índices 1.
  - ▶ Os números do conjunto de índices se referem a diferentes conjuntos de pares de índices i5 e i7. O Conjunto 1 e o Conjunto 2 fornecem, ambos, a diversidade de índice. Dois conjuntos de índices são oferecidos para ajudar a impedir o esgotamento de um único conjunto.
  - ▶ Escolha o número de amostras mais próximo do número de amostras que você está testando. Se o número exato de amostras não estiver na lista, selecione o número mais próximo, mas inferior ao número que você está testando. Por exemplo, se você quiser testar 18 amostras, selecione 16 amostras.
  - ▶ Poços de amostras sugeridos e combinações de índices que atendem aos requisitos de diversidade de índice são destacados em verde.

### Importar arquivos de manifesto para a execução

- 1 Certifique-se de que os manifestos que você quer importar estejam disponíveis em um local acessível da rede ou em uma unidade USB.
- 2 Selecione **Import Manifests** (Importar manifesto).
- 3 Navegue até o arquivo de manifesto e selecione os manifestos que você deseja adicionar.


**OBSERVAÇÃO** Para disponibilizar os arquivos de manifesto para todas as execuções usando o módulo de análise de variante de linha genética ou somática, adicione manifestos usando o recurso **Module Settings** (Configurações do módulo). Esse recurso requer permissões em nível de usuário admin. Para obter mais informações, consulte o *Guia de referência do instrumento NextSeq 550Dx* (documento n.º 1000000009513).

### Especificar amostras para a execução


Especifique amostras para a execução usando uma das opções e as orientações a seguir.

- ▶ **Informar amostras manualmente** — use a tabela em branco na tela Create Run (Criar execução).
- ▶ **Importar amostras** — navegue até um arquivo externo em um formato de valores separados por vírgulas (\*.csv). Há um modelo disponível para download na tela Create Run (Criar execução).

#### Insira amostras manualmente

- 1 Insira um nome de amostra exclusivo (**módulo de análise de variante somática**) ou um ID da amostra (**módulo de análise da variante de linha genética**).  
Use caracteres alfanuméricos, sublinhados ou travessões.
- 2 **[Opcional]** Para amostras de controle positivas ou negativas, clique com o botão direito do mouse e selecione o tipo de controle.  
O controle em um poço de amostras preenche automaticamente o poço correspondente no outro pool com o mesmo controle.
- 3 **[Opcional]** Informe uma descrição para a amostra no campo Sample Description (Descrição da amostra).  
Use caracteres alfanuméricos, sublinhados ou travessões.
- 4 Selecione um adaptador de Índice 1 na lista suspensa Index 1 (i7) (Índice 1 [i7]).  
Quando você usa poços de amostra sugeridos, o software preenche automaticamente os adaptadores de índice i7 e i5 que atendem aos requisitos de índice de diversidade. Se o número exato de amostras que você está testando não estiver na lista, selecione adaptadores de índice para poços extras.
- 5 Selecione um adaptador de Índice 2 na lista suspensa Index 2 (i5) (Índice 2 [i5]).
- 6 Selecione um arquivo de manifesto na lista suspensa Manifest (Manifesto).  
As amostras no Pool A exigem um manifesto diferente do das amostras no Pool B.
- 7 Escolha uma opção para visualizar, imprimir ou salvar o layout da placa como referência para preparar bibliotecas:
  - Selecione o ícone  **Print** (Imprimir) para exibir o layout da placa. Selecione **Print** (Imprimir) para imprimir o layout da placa.
  - Clique em **Export** (Exportar) para exportar informações da amostra para um arquivo externo.
- 8 Selecione **Save Run** (Salvar execução).

#### Importar amostras

- 1 Selecione **Import Samples** (Importar amostras) e navegue até o local do arquivo de informações das amostras. Há dois tipos de arquivos que você pode importar.
  - Selecione **Template** (Modelo) na tela Create Run (Criar execução) para fazer um novo layout da placa. O arquivo modelo contém os cabeçalhos corretos das colunas para importação. Especifique as informações das amostras em cada coluna para as amostras na execução. Exclua as informações de exemplo em células não utilizadas e depois salve o arquivo.
  - Use um arquivo de informações das amostras que tenha sido exportado do módulo de variante de linha genética ou somática usando o recurso Export (Exportar).
- 2 Selecione o ícone  **Print** (Imprimir) para exibir o layout da placa.
- 3 Selecione **Print** (Imprimir) para imprimir o layout da placa como referência para preparar bibliotecas.
- 4 Selecione **Save Run** (Salvar execução).

## Preparar o cartucho de reagente

Siga sempre as orientações do cartucho de reagente com cuidado para fazer um sequenciamento bem-sucedido.

- 1 Remova o cartucho de reagente do armazenamento de -15 °C a -25 °C.
- 2 Escolha um dos seguintes métodos para descongelar os reagentes. Não mergulhe o cartucho. Depois que o cartucho for descongelado, seque-o antes de ir para a próxima etapa.

Temperatura	Tempo de descongelamento	Limite de estabilidade
Banho-maria de 15 °C a 30 °C	60 minutos	Não ultrapassar 6 horas
2 °C a 8 °C	7 horas	Não ultrapassar 5 dias

**OBSERVAÇÃO** Se mais de um cartucho for descongelado no mesmo banho-maria, deixe mais tempo para descongelar.

- 3 Inverta o cartucho cinco vezes para misturar os reagentes.
- 4 Inspeção a parte inferior do cartucho para garantir que os reagentes estejam descongelados e livres de precipitados. Confirme se as posições 29, 30, 31 e 32 estão descongeladas porque elas são as maiores e levam mais tempo para descongelar.
- 5 Bata levemente na bancada para reduzir as bolhas de ar.  
Para obter melhores resultados, prossiga diretamente para o carregamento da amostra e a configuração da execução.

## Preparar a lâmina de fluxo

- 1 Retire uma caixa nova de lâminas de fluxo do armazenamento de 2 °C a 8 °C.
- 2 Remova a embalagem metálica da caixa e separe à temperatura ambiente durante 30 minutos.

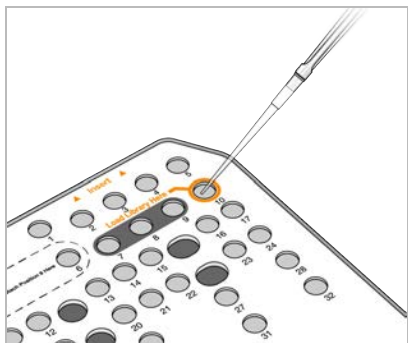
## Preparar bibliotecas para sequenciamento

Desnatura e dilua suas bibliotecas até um volume de carregamento de 1,3 ml. Na prática, a concentração de carga pode variar dependendo dos métodos de preparação e de quantificação da biblioteca. A diluição de bibliotecas de amostras depende da complexidade dos pools de oligonucleotídeos. Para obter instruções sobre o preparo de bibliotecas de amostras para o sequenciamento, inclusive diluição e pooling de bibliotecas, consulte a seção Instruções de uso quanto ao kit de preparação de bibliotecas aplicável. É necessária a otimização da densidade de cluster no NextSeq 550Dx.

## Carregar bibliotecas no cartucho de reagente

- 1 Limpe o selo de alumínio que cobre o reservatório n.º 10 rotulado **Load Library Here** (Carregar biblioteca aqui) usando um lenço com poucos fiapos.
- 2 Perfure o selo com a ponta de uma pipeta limpa de 1 ml.
- 3 Carregue 1,3 ml de bibliotecas preparadas no reservatório n.º 10 rotulado **Load Library Here** (Carregar biblioteca aqui). Evite tocar no selo de alumínio enquanto as bibliotecas são colocadas.

Figura 1 Carregar bibliotecas



## Configurar uma execução do sequenciamento

- 1 Faça login no NextSeq 550Dx com sua senha do software Local Run Manager.



- 2 Na tela Home (Início) do software NOS, selecione **Sequence** (Sequenciamento).
- 3 Selecione uma execução na lista e, em seguida, selecione **Next** (Avançar).

Uma série de telas de configuração de execução é aberta na seguinte ordem: Load Flow Cell (Carregar lâmina de fluxo), Load Buffer Cartridge (Carregar cartucho de solução tampão), Load Reagent Cartridge (Carregar cartucho de reagente) e Pre-run Check (Verificação pré-execução).
- 4 Quando a tela Load Flow Cell (Carregar lâmina de fluxo) for exibida, limpe e, em seguida, carregue a lâmina de fluxo.
  - ▶ Remova a lâmina de fluxo da embalagem metálica.
  - ▶ Abra o pacote plástico transparente e remova a lâmina de fluxo.
  - ▶ Limpe a superfície de vidro da lâmina de fluxo com um pano sem fiapos com álcool. Seque o vidro com um lenço para laboratório com poucos fiapos.
  - ▶ Certifique-se de que a superfície de vidro da lâmina de fluxo esteja limpa. Se necessário, repita a etapa de limpeza.
  - ▶ Remova a lâmina de fluxo usada em uma execução anterior.
  - ▶ Alinhe a lâmina de fluxo sobre os pinos de alinhamento e posicione-a no estágio da lâmina de fluxo.
- 5 Selecione **Load** (Carregar).

A porta se fecha automaticamente, a identificação da lâmina de fluxo é exibida na tela e os sensores são verificados.
- 6 Siga os avisos do software para esvaziar o receptáculo de reagentes usado, carregue o cartucho de solução tampão NextSeq 550Dx e o cartucho de reagente NextSeq 550Dx.

Quando a solução tampão e os cartuchos de reagente NextSeq 550Dx são carregados, o software lê e registra o rastreamento por identificador por radiofrequência (RFID). Os IDs da solução tampão e do cartucho de reagente aparecem na tela e os sensores são verificados.
- 7 Quando a verificação automática antes da execução for concluída, selecione **Start** (Iniciar) (não será necessário se for configurada para iniciar automaticamente).
- 8 A tela Sequencing (Sequenciamento) abre quando a execução começa. Essa tela fornece uma representação visual da execução em andamento, incluindo intensidades e pontuações de qualidade (Q-scores).

## Resultados

O Real-Time Analysis (RTA) é um software integrado que faz a análise da imagem e a identificação de bases e atribui uma pontuação de qualidade a cada base de cada ciclo de sequenciamento. Quando a análise primária é concluída, o módulo do Local Run Manager selecionado no NextSeq 550Dx Instrument inicia a análise secundária automaticamente. Os processos de análise secundária descritos aqui são para os módulos de variantes de linha genética e somáticas.

## Demultiplexação

A demultiplexação compara cada sequência da leitura de índice com as sequências de índice especificadas para a execução. Nenhum valor de qualidade é considerado nesta etapa.

As leituras de índice são identificadas com o uso das seguintes etapas:

- ▶ As amostras são numeradas a partir de 1 com base na ordem em que são listadas para a execução.
- ▶ A amostra número 0 é reservada para clusters que não foram atribuídos a uma amostra.
- ▶ Os clusters são atribuídos a uma amostra quando a sequência de índice corresponde exatamente ou quando há no máximo uma só disparidade por leitura de índice.

## Geração do arquivo FASTQ

Depois da demultiplexação, o software gera arquivos de análise intermediários no formato FASTQ, que é um formato de texto usado para representar sequências. Os arquivos FASTQ contêm leituras de cada amostra e as pontuações de qualidade associadas. Os clusters que não passaram pelo filtro são excluídos.

Cada arquivo FASTQ contém leituras de apenas uma amostra e o nome dessa amostra é incluído no nome do arquivo FASTQ. Nos módulos de variantes de linha genética e somática, oito arquivos FASTQ são gerados por amostra por pool de oligonucleotídeos, quatro da Leitura 1 e quatro da Leitura 2. Esse resultado gera um total de 8 e 16 arquivos FASTQ por amostra para variantes de linha genética e somática, respectivamente. Os arquivos FASTQ são a entrada primária para alinhamento.

## Alinhamento

Durante a etapa de alinhamento, o algoritmo Smith-Waterman escalonado alinha clusters de cada amostra em relação a sequências de Amplicon especificadas no arquivo de manifesto.

O algoritmo Smith-Waterman escalonado executa alinhamentos de sequência semiglobal para determinar regiões similares entre duas sequências. Em vez de comparar a sequência total, o algoritmo Smith-Waterman compara segmentos de todos os tamanhos possíveis.

Cada leitura tipo paired-end é avaliada em termos de seu alinhamento com as sequências de sondagem relevantes para essa leitura.

- ▶ A Leitura 1 é avaliada em relação ao complemento reverso dos oligos específicos de lugar a jusante (DLSO, Downstream Locus-Specific Oligos).
- ▶ A Leitura 2 é avaliada em relação aos oligos específicos de lugar a montante (ULSO, Upstream Locus-Specific Oligos).
- ▶ Se o início de uma leitura corresponder a uma sequência de sondagem com no máximo uma diferença, o tamanho total da leitura estará alinhado em relação ao destino do amplicon para essa sequência.
- ▶ Se o início de uma leitura corresponder a uma sequência de sondagem com no máximo três diferenças (disparidades ou alterações devido a indels de alinhamento), o tamanho total da leitura será alinhado em relação ao destino do Amplicon para essa sequência.
- ▶ Os indels dentro dos DLSO e dos ULSO não são observados tendo em conta a química do ensaio.

Os alinhamentos são filtrados dos resultados de alinhamento com base nas taxas de disparidade em relação à região de interesse ou ao Amplicon completo, dependendo do tamanho do Amplicon. Os alinhamentos filtrados são gravados em arquivos de alinhamento como desalinhados e não são usados em identificação de variantes.

## Identificação da variante

O identificador de variantes Pisces faz identificações de variantes de nucleotídeo único (SNV, single nucleotide variant) e de indels nas bibliotecas preparadas para o instrumento.

## Relatórios e arquivos de saída adicionais

Os módulos de análise de variantes produzem relatórios em PDF e delimitados por tabulação (\*.txt) que mostram métricas, como profundidade do sequenciamento e contagens de variantes. Os módulos também produzem arquivos de saída, como arquivos em VCF (Variant Call Format, Formato de identificação de variante) e em formato de identificação de variante genômica (gVCF, genome Variant Call Format) para aplicações de identificação de variante.

## Procedimentos de controle de qualidade

O software do NextSeq 550Dx avalia cada execução, amostra e identificação de bases com relação a métricas de controle de qualidade. Também são recomendados controles positivos e negativos na preparação das bibliotecas e que precisam ser avaliados. Avalie os controles como mostrado a seguir:

- **Controle negativo (controle sem modelo) ou outro controle negativo**— deve gerar o resultado esperado. Se o controle negativo gerar um resultado diferente do esperado, pode ter ocorrido erro no rastreamento das amostras, registro incorreto de primers de índice ou contaminação.
- **Amostra de controle positivo**— deve gerar o resultado esperado. Se o controle positivo gerar um resultado diferente do esperado, pode ter ocorrido erro no rastreamento das amostras, registro incorreto de primers de índice ou contaminação.

## Características de desempenho

As características de desempenho do NextSeq 550Dx Instrument foram estabelecidas usando os módulos de variante somática e de linha genética com o TruSeq Custom Amplicon Kit Dx e o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e confirmadas usando o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Os estudos incluíram indexação de amostras, transferência de amostras, entrada de DNA, sensibilidade analítica (limite de branco/ limite de detecção), exatidão, precisão, comparação de métodos e reprodutibilidade. Os estudos analíticos usando o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) foram destinados a avaliar os dados de desempenho previamente estabelecidos com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Os resultados demonstram que os kits de reagentes (v2 e v2.5) têm desempenho comparável usando o TruSeq Custom Amplicon Kit Dx. Consulte o *folheto informativo do TruSeq Custom Amplicon Kit Dx* quanto às características de desempenho relacionadas aos fatores pré-analíticos, como métodos de extração ou substâncias interferentes.

## Definições de cálculos usados nas características de desempenho

- 1 O percentual de concordância positiva (PPA, Positive Percent Agreement) é calculado como a proporção de locais classificados como variantes por um método de referência que são corretamente relatados pelo ensaio.
  - ▶  $(\text{N.º de locais de variantes corretamente relatados pelo ensaio}) / (\text{N.º total de locais de variantes})$   
Os locais de variantes relatados pelo ensaio que são concordantes com o método de referência são positivos verdadeiros (TPs, True Positives). Os locais de variantes relatados como identificações de referência ou como identificações de variantes diferentes pelo ensaio são falsos negativos (FNs, False Negatives).
- 2 O percentual de concordância negativa (NPA, Negative Percent Agreement) é calculado como a proporção de locais classificados como tipo selvagem por um método de referência que são corretamente relatados pelo ensaio.
  - ▶  $(\text{N.º de locais do tipo selvagem corretamente relatados pelo ensaio}) / (\text{N.º total de locais do tipo selvagem})$   
Os locais do tipo selvagem relatados pelo ensaio que são concordantes com o método de referência são negativos verdadeiros (TNs, True Negatives). Os locais do tipo selvagem relatados como variantes pelo ensaio são falsos positivos (FPs).
- 3 O percentual geral de concordância (OPA, Overall Percent Agreement) é calculado como a proporção de locais corretamente relatados pelo ensaio, relativos a um método de referência.
  - ▶  $([\text{N.º de locais de variantes corretamente relatados pelo ensaio}] + [\text{N.º de locais do tipo selvagem corretamente relatados pelo ensaio}]) / ([\text{N.º total de locais de variantes}] + [\text{N.º total de locais do tipo selvagem}])$
- 4 Os cálculos de PPA, NPA e OPA não incluem não identificações (locais de variantes ou de referência que não atendem a um ou mais filtros de qualidade).
- 5 A taxa de identificação de autossomo é calculada como o número total de locais que passam pelos filtros dividido pelo número total de posições sequenciadas para os cromossomos 1–22; os cromossomos X e Y são excluídos. Essa medida não considera a concordância das identificações com o método de referência.

## Desempenho do NextSeq 550Dx High Output Reagent Kit v2 (300 Cycles)

### Indexação de amostras

Os primers de índice de amostras, adicionados durante a preparação das bibliotecas, atribuem uma sequência única a cada amostra de DNA. Essas sequências únicas possibilitam que várias amostras sejam agrupadas em uma única execução do sequenciamento. A indexação de amostras é usada para os fluxos de trabalho de linha genética e somático. O objetivo deste estudo foi estabelecer o número mínimo (8) e máximo (96) de amostras que podem ser processadas em uma única execução do sequenciamento pelo NextSeq 550Dx Instrument. Foram testadas oito amostras exclusivas de Platinum Genome com 12 combinações diferentes de primers de indexação por amostra. Os resultados da amostra de quatro execuções do sequenciamento com o uso do módulo de variante de linha genética foram comparados com os Platinum Genomes versão 2016-1.0.

Para o primeiro conjunto de execuções, foram testadas 96 bibliotecas de amostras indexadas exclusivamente com um ensaio representativo destinado a consultar uma variedade de genes, abrangendo 12.588 bases por fita em todos os 23 cromossomos humanos, para verificar a capacidade do ensaio de fazer consistentemente uma identificação de genótipos de uma determinada amostra em combinações diferentes de primers de indexação. Para o segundo conjunto de execuções, foram sequenciadas oito bibliotecas de amostras exclusivamente indexadas em duas execuções do sequenciamento para verificar o número mínimo de índices compatíveis.

Para as 96 execuções de índices, o PPA de SNVs variou de 98,7% a 100%, o PPA de inserções e deleções foi de 100% e o NPA foi de 100% para cada uma das 96 combinações de índices. As 8 execuções de índices apresentaram valores de PPA de 100% (SNVs, inserções e deleções) e de NPA de 100% para cada uma das oito combinações de índices.

### Transferência de amostra

O NextSeq 550Dx Instrument permite o sequenciamento de várias amostras e de controles em uma única execução do sequenciamento. Foi realizado um estudo para avaliar a extensão da transferência de amostras em uma execução do sequenciamento (dentro da execução) e entre execuções do sequenciamento (de execução para execução). Duas amostras de Platinum Genome, uma do sexo masculino e uma do sexo feminino, foram testadas em um ensaio representativo destinado a consultar vários genes, abrangendo 12.588 bases (150 amplicons) em 23 cromossomos diferentes, incluindo cromossomos de ambos os sexos. As bibliotecas foram sequenciadas no NextSeq 550Dx Instrument usando o módulo de variante de linha genética. A transferência de amostras do sexo masculino em amostras do sexo feminino foi observada pela presença de leituras de amplicon de cromossomo Y em amostras do sexo feminino.

A transferência dentro da execução pode ser introduzida durante a clusterização, a identificação de bases de ciclo de índice e a demultiplexação de amostras. Para testar a transferência de amostras dentro de uma execução do sequenciamento, um pool de bibliotecas, consistindo em 46 réplicas, cada uma delas com amostras do sexo masculino e do sexo feminino e quatro controles sem modelo, foi sequenciado uma vez no NextSeq 550Dx Instrument. A transferência de amostras dentro da execução foi avaliada pela comparação da cobertura do amplicon do cromossomo Y de cada réplica do sexo feminino com a cobertura média do amplicon do cromossomo Y de todas as réplicas do sexo masculino do pool. A mediana observada na transferência dentro da execução foi de 0,084%.

Para testar a transferência de amostras de execução para execução, foram preparados e sequenciados consecutivamente dois pools de bibliotecas em um NextSeq 550Dx Instrument. O primeiro pool continha 46 réplicas de amostra do sexo feminino e dois controles sem modelo. O segundo pool continha 46 réplicas de amostra do sexo masculino e dois controles sem modelo. Ambos os pools usaram o mesmo conjunto de adaptadores de índice. O pool do sexo feminino foi sequenciado primeiro, seguido de uma execução subsequente do sequenciamento com o pool do sexo masculino, seguido de outra execução do sequenciamento repetida do pool do sexo feminino. A transferência de amostras de execução para execução foi avaliada pela comparação da cobertura do amplicon do cromossomo Y entre as réplicas correspondentes da execução de repetição do pool do sexo feminino e da execução do pool do sexo masculino. A mediana observada na transferência de execução para execução foi de 0,0076%.

## Entrada de DNA

Sangue (linha genética)

O intervalo de entrada de DNA do sangue para preparação de bibliotecas com o TruSeq Custom Amplicon Kit Dx usando o fluxo de trabalho do módulo de variante de linha genética foi estabelecido para o NextSeq 550Dx Instrument. Esse intervalo foi avaliado com a realização de um estudo de diluição em série com o uso de 13 amostras de Platinum Genomes com um ensaio representativo criado para consultar vários genes, cobrindo 12.588 bases em 23 cromossomos diferentes. A biblioteca foi sequenciada em dois instrumentos NextSeq 550Dx usando um lote do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Cinco amostras foram testadas em duplicata em cinco níveis de entrada de DNA que variaram de 250 ng a 12 ng (250 ng, 100 ng, 50 ng, 25 ng e 12 ng). Oito amostras foram testadas como uma réplica simples em cada um dos cinco níveis de entrada de DNA. Para a determinação da exatidão, os genótipos da amostra foram comparados com Platinum Genomes versão 2016-1.0. Os resultados foram determinados para cada nível de entrada. O PPA de cada tipo de variante (SNVs, inserções e deleções) é apresentado na [Tabela 1](#); o NPA é apresentado na [Tabela 2](#). Todos os níveis de entrada apresentaram exatidão semelhante. A entrada de DNA recomendada para o TruSeq Custom Amplicon Kit Dx é de 50 ng, sendo que 25 ng e 100 ng representam os limites inferior e superior para atender às características de desempenho.

Tabela 1 Resultados de PPA de cada entrada de DNA por tipo de variante

Entrada de DNA (ng)	Tipo de variante	Variante prováveis	TP	FN	Não identificações de variante	PPA (%)
12	SNV	2412	2381	31	0	98,7
25			2404	8	0	99,7
50			2403	9	0	99,6
100			2412	0	0	100
250			2412	0	0	100
12	Inserção	808	784	3	21	99,6
25			781	5	22	99,4
50			786	2	20	99,8
100			786	0	22	100
250			786	0	22	100
12	Deleção	758	732	12	14	98,4
25			737	7	14	99,1
50			742	2	14	99,7
100			744	0	14	100
250			744	0	14	100

Tabela 2 NPA para cada entrada de DNA

Entrada de DNA (ng)	TN	FP	Não identificações de referência	NPA (%)
12	430940	4	26	>99,9
25	430936	0	34	100
50	430936	2	32	>99,9
100	430942	0	28	100
250	430942	0	28	100

## FFPE (somático)

Foi estabelecido o intervalo de entrada de DNA fixado em formalina e embebido em parafina (FFPE, Formalin-Fixed Paraffin-Embedded) para a preparação de biblioteca do TruSeq Custom Amplicon Kit Dx com o fluxo de trabalho do módulo de variante somática para o NextSeq 550Dx Instrument. O intervalo de entrada de DNA foi avaliado por um estudo de diluição em série com o uso de três amostras de Platinum Genomes com um ensaio representativo criado para consultar vários genes, abrangendo 12.588 bases em 23 cromossomos diferentes. As linhagens celulares GM12878 e GM12877 do Platinum Genome foram fixadas em formalina e embebidas em parafina seguidas pela extração de DNA. A linhagem GM12878 foi diluída com a linhagem GM12877 de modo que as frequências alélicas (VAFs) de 81 variantes (55 SNVs, 10 inserções e 16 deleções) ficaram próximas de 0,025, 0,05 ou 0,10. Além disso, cada amostra tinha 91 variantes com frequências mais elevadas (VAF) de até 1,0. As amostras foram processadas em duplicata em cinco níveis de entrada de DNA com ciclos quantitativos delta (dCq, delta quantitative cycle) médios de 2,1, 3,6, 4,6, 6,0 e 7,8 medidos pelo TruSeq Custom Amplicon Dx - FFPE QC Kit. Cada biblioteca foi sequenciada em dois instrumentos NextSeq 550Dx usando um lote do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Para determinação da exatidão, as identificações das variantes das amostras foram comparadas com Platinum Genomes versão 2016-1.0. O PPA para cada tipo de variante (SNVs, inserções e deleções) é apresentado na [Tabela 3](#); o NPA é apresentado na [Tabela 4](#). A entrada de DNA recomendada para variantes com VAF de 0,05 ou acima é  $dCq \leq 4$ , sendo que 4,6 é o limite inferior para o cumprimento das características de desempenho.

Tabela 3 Resultados de PPA para cada entrada de DNA por tipo de variante

dCq médio	Tipo de variante	Variantes prováveis	Não identificações prováveis	VAF de diluição-alvo					
				0,025		0,05		0,10	
				Não identificações de variante	PPA (%)	Não identificações de variante	PPA (%)	Não identificações de variante	PPA (%)
2,1	SNV	808	Não aplicável.	196	100	0	100	0	100
3,6				250	99,3	4	100	0	100
4,6				251	94,6	51	99,2	5	100
6,0				257	65,3	213	91,4	100	100
7,8				254	69,3	185	90,7	100	100
2,1	Inserção	264	8	66	96,5	8	100	8	100
3,6				62	97,0	8	100	8	100
4,6				48	96,3	21	100	8	100
6,0				40	80,4	47	98,2	24	95,8
7,8				57	87,0	56	96,2	31	100
2,1	Deleção	304	16	58	100	16	100	16	100
3,6				80	100	16	100	16	100
4,6				65	95,4	28	100	16	100
6,0				78	74,8	105	94,0	36	100
7,8				76	75,0	79	95,1	57	98,8

Tabela 4 NPA para cada entrada de DNA

dCq médio	Tipo selvagem provável	VAF de diluição-alvo					
		0,025		0,05		0,10	
		Não identificações de referência	NPA (%)	Não identificações de referência	NPA (%)	Não identificações de referência	NPA (%)
2,1	93688	344	100	260	100	324	100
3,6		400	100	332	100	380	100
4,6		1308	100	1336	100	784	100
6,0		3900	>99,9	3296	>99,9	2996	100
7,8		3020	>99,9	2880	>99,9	2448	>99,9

### Sensibilidade analítica (limite de branco [LoB] e limite de detecção [LoD])

Este estudo foi conduzido para avaliar o limite de branco (LoB) e o limite de detecção (LoD) para o módulo de variante somática no NextSeq 550Dx Instrument. O estudo foi realizado com um ensaio representativo criado para consultar vários genes, abrangendo 12.588 bases em 23 cromossomos diferentes. As linhagens celulares GM12878 e GM12877 do Platinum Genome foram fixadas em formalina e embebidas em parafina seguidas pela extração de DNA. A linhagem GM12878 foi diluída com a linhagem GM12877 de modo que as frequências de 74 variantes (53 SNVs, 7 inserções e 14 deleções) foram  $0,05 \pm 0,02$ . A linhagem GM12877 e a linhagem diluída GM12878 (GM12878-D) foram testadas durante seis dias de início consecutivos, alternando entre dois lotes do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles), para um total de seis execuções do sequenciamento. O teste resultou em 60 réplicas para cada variante em GM12878-D e 72 réplicas para cada coordenada do tipo selvagem correspondente em GM12877 para cada lote de reagente. O LoB e o LoD foram calculados usando a abordagem clássica indicada no documento CLSI EP17-A2 usando a opção não paramétrica. O LoB e o LoD foram calculados para SNVs, inserções e deleções separadamente com a criação de pools das frequências de variantes para um certo tipo de variante. O erro do tipo I foi definido como 0,01 e o erro do tipo II foi definido como 0,05.

Para o LoB, as frequências de variantes do pool foram classificadas da mais baixa para a mais alta, e a 99.<sup>a</sup> posição da classificação para cada lote de reagente e cada tipo de variante foi calculada (Tabela 5). O módulo de variante somática usa um corte (LoB eficaz) com VAF de 0,026 para determinar a detecção qualitativa das variantes. O LoB calculado verificou que esse corte resulta em um erro do tipo I de, no máximo, 0,01.

Tabela 5 Limite de branco

Tipo de variante	Total de observações	LoB do lote de reagente 1 (%)	LoB do lote de reagente 2 (%)
SNV	3816	0,77	0,77
Inserção	504	0,56	0,56
Deleção	1008	1,20	1,20

Para o LoD, o percentual da frequência de mutação individual para cada lote de reagente e tipo de variante abaixo do corte de 0,026 foi calculado na Tabela 6. Como os percentuais foram menores do que o erro do tipo II de 5% (0,05), a mediana das frequências de variantes combinadas foi calculada como o LoD (Tabela 6). O LoD para cada tipo de variante foi tomado como o maior dos dois valores calculados para os dois lotes de reagente: 4,97% para SNVs, 5,12% para inserções e 5,26% para deleções.

Tabela 6 Limite de detecção

Lote de reagente	Tipo de variante	Total de observações	N.º de medições de VAF <2,6%	% de medições de VAF <2,6%	Limite de detecção (%)
1	SNV	3180	53	1,7	4,94
	Inserção	420	6	1,4	5,08
	Deleção	840	7	0,8	5,22
2	SNV	3180	51	1,6	4,97
	Inserção	420	5	1,2	5,12
	Deleção	840	7	0,80	5,26

## Exatidão

### Linha genética

O estudo a seguir foi conduzido para avaliar a exatidão da identificação de variantes do módulo de variante de linha genética no NextSeq 550Dx Instrument usando o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Treze amostras de Platinum Genome foram testadas por meio de um ensaio representativo destinado a consultar uma variedade de genes abrangendo 12.588 bases (150 amplicons) em 23 diferentes cromossomos. Foram conduzidas nove execuções usando três instrumentos de sequenciamento, três lotes de reagentes e três operadores ao longo de cinco dias de execução. Foi determinada a exatidão para SNVs, inserções e deleções pela comparação dos resultados com um método de referência composto bem caracterizado, Platinum Genomes versão 2016-1.0. Foram definidas regiões genômicas de confiança com base nesse método de referência, a menos que especificado de outra forma.

Tabela 7 Resumo de concordância da linha genética

Crítérios	Total de observações <sup>1</sup>	Resultado por observação <sup>2</sup>	Resultado por execução <sup>3</sup>
PPA para SNV	819	98,7	>99,9
PPA para inserções	819	95,0	98,9
PPA para deleções	819	100	100
NPA	819	100	100
OPA	819	>99,9	>99,9

<sup>1</sup>Calculado como o número de amostras por execução (91) x número de execuções (9) = 819.

<sup>2</sup>Valor mais baixo observado por réplica de amostra em todas as 9 execuções.

<sup>3</sup>Valor mais baixo quando os dados de cada execução são analisados em conjunto.

A **Tabela 8** contém os dados do estudo apresentados com concordância percentual positiva e negativa para cada amostra, sendo que os resultados da variante são comparados com os Platinum Genomes versão 2016-1.0 para cálculos de PPA. Os três tipos de variante (SNVs, inserções e deleções) são combinados. Como o método de referência só fornece resultados para as variantes de nucleotídeo único e inserções/deleções, os resultados de base não variantes são comparados com sequências de referências do genoma humano versão hg19 para os cálculos do NPA.



Tabela 8 Concordância da linha genética por amostra

Amostra	Taxa de identificação média	Variantes prováveis <sup>1</sup>	TP	FN	Não identificações de variante	TN	FP	PPA	NPA	OPA
NA12877	>99,9	4788	4788	0	0	756762	0	100	100	100
NA12878	>99,9	8505	8379	1	125	751464	0	>99,9	100	>99,9
NA12879	>99,9	6048	5985	5	58	757701	0	99,9	100	>99,9
NA12880	>99,9	6993	6930	0	63	757638	0	100	100	100
NA12881	>99,9	7875	7811	3	61	751653	0	>99,9	100	>99,9
NA12882	>99,9	6300	6174	3	123	754803	0	>99,9	100	>99,9
NA12883	>99,9	7119	7056	0	63	751905	0	100	100	100
NA12884	>99,9	7182	7119	6	57	754146	0	99,9	100	>99,9
NA12885	>99,9	7686	7560	2	124	754173	0	>99,9	100	>99,9
NA12886	>99,9	7245	7182	7	56	752469	0	99,9	100	>99,9
NA12887	>99,9	7119	7119	0	0	750645	0	100	100	100
NA12888	>99,9	6804	6804	0	0	756065	0	100	100	100
NA12893	>99,9	7434	7371	1	62	750015	0	>99,9	100	>99,9

<sup>1</sup> Número total de variantes em todas as réplicas de amostras em 9 execuções.

A **Tabela 9** contém os dados do estudo apresentados para cada amostra, sendo que os resultados da variante são comparados com o método de referência composto bem caracterizado. A detecção é avaliada para cada tipo de variante: SNVs, inserções e deleções, separadamente. As posições de referência são excluídas.

Tabela 9 Concordância da linha genética por amostra por tipo de variante

>Amostra	SNVs			Inserções			Deleções		
	>Provável	>TP	>FN	>Provável	>TP	>FN	Provável	TP	FN
NA12877	2331	2331	0	1323	1323	0	1134	1134	0
NA12878	5733	5733	0	1260	1197	1	1512	1449	0
NA12879	3591	3591	0	1323	1260	5	1134	1134	0
NA12880	4221	4221	0	1512	1512	0	1260	1197	0
NA12881	4914	4913	1	1512	1449	2	1449	1449	0
NA12882	3717	3717	0	1386	1323	3	1197	1134	0
NA12883	4284	4284	0	1449	1449	0	1386	1323	0
NA12884	4284	4284	0	1575	1512	6	1323	1323	0
NA12885	4725	4725	0	1575	1512	2	1386	1323	0
NA12886	4347	4347	0	1449	1386	7	1449	1449	0
NA12887	4284	4284	0	1323	1323	0	1512	1512	0
NA12888	4158	4158	0	1449	1449	0	1197	1197	0
NA12893	4599	4599	0	1386	1323	1	1449	1449	0

As amostras foram analisadas novamente para a identificação de pequenas inserções e deleções (indels). Um resumo geral é apresentado na **Tabela 10**. Foi encontrado um total de 71 indels, com tamanho variando de 1–24 bp para inserções e 1–25 bp para deleções.

Tabela 10 Resumo da detecção de indels na linha genética

Tipo de variante	Variantes prováveis	TP	FN	Não identificações de variante	PPA
Inserção	18522	18018	27	477	99,9
Deleção	17388	17073	0	315	100

O ensaio representativo consistiu em 150 amplicons destinados a abranger conteúdo genômico diversificado. O conteúdo GC dos amplicons variou de 0,19 a 0,87. Os amplicons também tiveram uma faixa de nucleotídeos únicos (p. ex., PoliA, PoliT), dinucleotídeos e repetições de trinucleotídeos. Os dados foram comparados para cada amplicon (Tabela 11) para determinar o efeito do conteúdo genômico no percentual de identificações corretas. O percentual de identificações corretas consiste em identificações de variância e de referência e é inferior a 100% no caso de identificações incorretas ou não identificações.

Tabela 11 Exatidão no nível de amplicons da linha genética

Amplicon	Cromossomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
1	1	36450499	36450591	93	93	Indel	0,22	76167	0	0	100
2	1	109465122	109465200	79	79	PoliA (5), PoliC (5), indel	0,38	64701	0	0	100
3	1	218353867	218353957	91	91	Indel	0,4	74529	0	0	100
4	1	223906657	223906748	92	92	Indel	0,49	75348	0	0	100
5	1	228526602	228526682	81	81	PoliG (5)	0,69	66339	0	0	100
6	1	236372039	236372108	70	70	PoliT (10), indel	0,39	57330	0	0	100
7	1	247812041	247812128	88	88	PoliA (5), CT(3), TAA(3), indel	0,27	72072	0	0	100
8	2	55862774	55862863	90	90	Indel	0,28	73710	0	0	100
9	2	87003930	87004009	80	80	Indel	0,38	65520	0	0	100
10	2	177016721	177016805	85	81	N/A	0,65	66339	0	0	100
11	2	186625727	186625801	75	75	PoliA (8)	0,35	61425	0	0	100
12	2	190323504	190323591	88	88	PoliT (5)	0,42	72072	0	0	100
13	2	200796740	200796826	87	87	PoliT (5), indel	0,31	71253	0	0	100
14	2	212245049	212245139	91	91	PoliT (5), PoliA (6), indel	0,3	74529	0	0	100
15	2	228147052	228147144	93	93	Indel	0,43	76167	0	0	100
16	2	235016350	235016422	73	73	PoliT (5), indel	0,42	59787	0	0	100
17	3	4466229	4466321	93	93	AT(3), indel	0,27	74823	0	1344	98,2
18	3	46620561	46620643	83	83	N/A	0,43	67977	0	0	100
19	3	49851331	49851400	70	70	CT(3), indel	0,49	57330	0	0	100

Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
20	3	189713161	189713248	88	88	PoliA (5), PoliT (5), PoliA (9), TG (3)	0,41	72072	0	0	100
21	3	190106030	190106104	75	74	Indel	0,57	60543	0	63	99,9
22	4	2233667	2233744	78	78	PoliA (6)	0,26	63882	0	0	100
23	4	7780541	7780637	97	97	PoliG (6), PoliT (5), PoliA (5)	0,42	79443	0	0	100
24	4	15688604	15688681	78	78	N/A	0,29	63882	0	0	100
25	4	56236521	56236586	66	62	PoliA (5), indel	0,36	50778	0	0	100
26	4	102839244	102839314	71	69	PoliA (5)	0,46	56511	0	0	100
27	4	164446743	164446804	62	62	PoliA (7), indel	0,27	50778	0	0	100
28	5	1882081	1882158	78	75	N/A	0,78	61425	0	0	100
29	5	14769061	14769144	84	84	GT(3), CCA(3)	0,62	68796	0	0	100
30	5	41069808	41069871	64	64	N/A	0,39	52416	0	0	100
31	5	74077114	74077196	83	83	PoliA (6), indel	0,3	67977	0	0	100
32	5	147475343	147475409	67	67	PoliT (5)	0,37	54873	0	0	100
33	5	149323731	149323821	91	91	CT(4), AG(3)	0,55	74529	0	0	100
34	5	155662213	155662287	75	75	Indel	0,43	61425	0	0	100
35	6	6318713	6318814	102	102	PoliG (6)	0,68	83538	0	0	100
36	6	24949983	24950074	92	92	Indel	0,63	75348	0	0	100
37	6	31084900	31084999	100	94	GCT(5), indel	0,61	76608	0	378	99,5
38	6	32147987	32148084	98	98	PoliT (5), TCT (3), CTT(3)	0,55	80262	0	0	100
39	6	32986864	32986958	95	95	Indel	0,53	77805	0	0	100
40	6	33408498	33408583	86	86	PoliC (6)	0,7	70434	0	0	100
41	6	41647401	41647495	95	94	PoliG (5), indel	0,61	76986	0	0	100
42	6	112435865	112435955	91	91	PoliA (5)	0,44	74529	0	0	100

Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
43	7	22202076	22202148	73	73	N/A	0,44	59787	0	0	100
44	7	66276100	66276187	88	88	Indel	0,35	72072	0	0	100
45	7	77365735	77365821	87	87	PoliA (7), AG(4)	0,26	71253	0	0	100
46	7	110939946	110940030	85	85	Indel	0,38	69615	0	0	100
47	7	128533468	128533557	90	90	PoliG (5), indel	0,62	73710	0	0	100
48	7	149503875	149503965	91	91	PoliG (6), PoliC (6), indel	0,71	74529	0	0	100
49	7	154404519	154404599	81	66	N/A	0,31	54054	0	0	100
50	7	156476507	156476599	93	93	Indel	0,35	76167	0	0	100
51	8	1817312	1817394	83	83	N/A	0,42	67977	0	0	100
52	8	24811020	24811109	90	89	PoliG (7), CTC (4), indel	0,61	72171	0	720	99,0
53	8	76518625	76518691	67	67	Indel	0,3	54873	0	0	100
54	9	103054909	103055006	98	98	PoliG (6)	0,67	80262	0	0	100
55	9	105586150	105586214	65	65	Indel	0,32	53235	0	0	100
56	9	107620823	107620918	96	96	N/A	0,49	78624	0	0	100
57	9	123769149	123769231	83	83	AT(3)	0,37	67977	0	0	100
58	9	138995345	138995441	97	97	PoliC (6), indel	0,68	79443	0	0	100
59	10	5987120	5987198	79	78	PoliG (5), indel	0,47	63882	0	0	100
60	10	11784629	11784726	98	91	GC(3)	0,87	74529	0	0	100
61	10	27317777	27317855	79	79	PoliT (5)	0,3	64701	0	0	100
62	10	33018351	33018440	90	90	PoliA (5), PoliT (5)	0,2	73710	0	0	100
63	10	45084159	45084253	95	95	Indel	0,35	77805	0	0	100
64	10	55892599	55892687	89	88	AC(11), indel	0,42	71747	0	325	99,5
65	10	101611250	101611329	80	80	N/A	0,49	65520	0	0	100
66	10	118351373	118351453	81	81	N/A	0,51	66339	0	0	100

Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
67	11	8159816	8159912	97	96	N/A	0,45	78624	0	0	100
68	11	30177648	30177717	70	70	Indel	0,46	57330	0	0	100
69	11	47470345	47470444	100	100	N/A	0,65	81900	0	0	100
70	11	59837679	59837740	62	62	Indel	0,37	50778	0	0	100
71	11	64418856	64418957	102	102	N/A	0,59	83538	0	0	100
72	11	93529612	93529684	73	73	PoliA (5)	0,4	59787	0	0	100
73	11	101347052	101347136	85	85	N/A	0,42	69615	0	0	100
74	11	102477336	102477426	91	91	PoliG (6)	0,55	74529	0	0	100
75	11	118406285	118406369	85	85	Indel	0,53	69615	0	0	100
76	11	120357801	120357885	85	85	PoliA (5), CA(3), indel	0,34	69615	0	0	100
77	11	125769313	125769397	85	85	GA(3)	0,52	69615	0	0	100
78	12	2834770	2834853	84	84	PoliC (5), indel	0,52	68796	0	0	100
79	12	26811004	26811096	93	93	PoliA (7), AC(4)	0,33	76167	0	0	100
80	12	30881766	30881846	81	81	N/A	0,49	66339	0	0	100
81	12	88474105	88474175	71	71	PoliA (6)	0,35	58149	0	0	100
82	12	120966872	120966966	95	95	PoliG (5)	0,68	77805	0	0	100
83	13	24167504	24167576	73	73	N/A	0,52	59787	0	0	100
84	13	25816961	25817049	89	88	PoliA (5), PoliT (7), PoliA (7), indel	0,22	72072	0	0	100
85	13	44880112	44880200	89	89	Indel	0,49	72891	0	0	100
86	13	77665218	77665294	77	77	Indel	0,39	63063	0	0	100
87	14	31619327	31619393	67	67	GA(3),TA(3)	0,39	54873	0	0	100
88	14	39517884	39517966	83	83	N/A	0,25	67977	0	0	100
89	14	46958962	46959034	73	72	PoliT (5), indel	0,19	58642	0	326	99,4
90	14	58050030	58050110	81	81	Indel	0,38	66339	0	0	100

Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
91	14	82390559	82390649	91	91	Indel	0,35	74529	0	0	100
92	14	92549544	92549609	66	66	PoliA (5)	0,41	54054	0	0	100
93	14	102808496	102808589	94	94	Indel	0,62	76986	0	0	100
94	15	43170751	43170848	98	96	PoliC (5)	0,45	78624	0	0	100
95	15	63446149	63446216	68	68	Indel	0,25	55692	0	0	100
96	15	77879807	77879901	95	93	PoliG (5), indel	0,68	76167	0	0	100
97	15	81625334	81625428	95	95	PoliT (6)	0,43	77805	0	0	100
98	15	85438263	85438334	72	71	Indel	0,65	58149	0	0	100
99	15	89817413	89817503	91	91	N/A	0,36	74529	0	0	100
100	15	89864274	89864343	70	70	Indel	0,56	57330	0	0	100
101	16	1894910	1894972	63	63	N/A	0,27	51597	0	0	100
102	16	28997904	28997998	95	95	PoliC (5)	0,67	77805	0	0	100
103	16	53682908	53682994	87	87	TA(3)	0,41	71253	0	0	100
104	16	57954406	57954509	104	104	PoliC (5)	0,67	85176	0	0	100
105	16	85706375	85706465	91	91	PoliT (5), indel	0,37	74529	0	0	100
106	17	3563920	3564008	89	89	GC(3)	0,64	72891	0	0	100
107	17	3594191	3594277	87	87	PoliC (5), indel	0,67	71247	0	6	100
108	17	3970090	3970180	91	91	Indel	0,46	74529	0	0	100
109	17	16084945	16085037	93	93	Indel	0,26	76167	0	0	100
110	17	33998759	33998849	91	89	PoliT (5)	0,54	72891	0	0	100
111	17	39589691	39589774	84	82	PoliA (13), indel (x2)	0,29	66343	27	788	98,8
112	17	41244394	41244484	91	91	PoliA (5)	0,34	74529	0	0	100
113	17	45438866	45438957	92	92	PoliA (7), AT(3), AT(4), AT(4), indel	0,26	75348	0	0	100
114	17	61502432	61502510	79	79	Indel	0,41	64413	0	288	99,6

Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
115	17	64023582	64023667	86	86	PoliT (7)	0,22	70434	0	0	100
116	17	72308237	72308320	84	84	GAG(3)	0,62	68796	0	0	100
117	18	2616456	2616522	67	67	GA(3)	0,31	54873	0	0	100
118	18	6980478	6980568	91	91	N/A	0,37	74529	0	0	100
119	18	9888026	9888094	69	69	PoliA (6), TG(3)	0,43	56511	0	0	100
120	18	38836999	38837073	75	75	PoliA (5), indel	0,37	61425	0	0	100
121	18	47405382	47405462	81	81	CTC(3), indel	0,47	66339	0	0	100
122	18	54815665	54815749	85	85	CT(3), indel	0,45	69615	0	0	100
123	18	59773996	59774060	65	65	N/A	0,48	53235	0	0	100
124	19	625143	625241	99	99	N/A	0,59	81081	0	0	100
125	19	18121418	18121491	74	74	N/A	0,68	60605	1	0	100
126	19	18186574	18186643	70	70	N/A	0,64	57330	0	0	100
127	20	746056	746149	94	94	N/A	0,61	76986	0	0	100
128	20	10633195	10633276	82	82	AC(3)	0,59	67158	0	0	100
129	20	17705633	17705708	76	76	CT(3)	0,58	62244	0	0	100
130	20	21766821	21766890	70	70	GT(3),TG(4), indel	0,46	57330	0	0	100
131	20	25278421	25278521	101	101	Indel	0,63	82719	0	0	100
132	20	50897302	50897368	67	67	Indel	0,36	54873	0	0	100
133	20	62331904	62331994	91	88	PoliG (6)	0,73	72072	0	0	100
134	20	62690860	62690946	87	87	Indel	0,57	71253	0	0	100
135	21	30300823	30300888	66	66	Indel	0,35	54054	0	0	100
136	21	33694176	33694273	98	98	PoliT (6), CA(3)	0,54	80262	0	0	100
137	21	36710706	36710792	87	87	GT(3), indel	0,39	71253	0	0	100
138	21	46644924	46644992	69	69	PoliA (6), AG(3), indel	0,32	56439	0	72	99,9



Amplicon	Cromo- somo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identifica- ções corretas	Identifica- ções incorretas	Não identifica- ções	% de identifica- ções corretas
139	21	46705575	46705664	90	90	PoliT (5), PoliA (6)	0,5	73710	0	0	100
140	22	25750774	25750873	100	100	Indel	0,63	81900	0	0	100
141	22	32439233	32439329	97	97	N/A	0,68	79443	0	0	100
142	22	37409844	37409940	97	97	Indel	0,46	79443	0	0	100
143	22	37637596	37637694	99	99	N/A	0,6	81081	0	0	100
144	22	47081347	47081438	92	92	Indel	0,66	75348	0	0	100
145	X	15870424	15870492	69	69	PoliT (5)	0,26	56511	0	0	100
146	X	135288543	135288611	69	69	PoliC (5)	0,62	56511	0	0	100
147	X	135290777	135290847	71	71	N/A	0,52	58149	0	0	100
148	Y	2655397	2655461	65	0	N/A	0,55	0	0	0	N/A
149	Y	2655519	2655609	91	0	N/A	0,48	0	0	0	N/A
150	Y	2655609	2655679	71	0	PoliA (5)	0,37	0	0	0	N/A

Os resultados do sequenciamento da amostra NA12878 foram comparados com um genótipo de alta confiança para NA12878, estabelecido pelo National Institutes of Standards and Technology (NIST) (v.2.19). Entre os 150 amplicons, 92 estavam totalmente contidos nas regiões genômicas de alta confiança, 41 amplicons tinham sobreposição parcial e 17 amplicons não tinham sobreposição na sequência NIST. Esse resultado acarretou 10.000 coordenadas por réplica para comparação. Identificações de base não variantes foram comparadas com sequências de referências do genoma humano versão hg19. Os resultados da exatidão são mostrados na **Tabela 12**.

Tabela 12 Concordância da linha genética da amostra NA12878 com o banco de dados do NIST

Amostra	N.º de amplicons	Taxa de identificação média	TP	FN	TN	FP	PPA	NPA	OPA
NA12878	133	>99,9	6552	1	610470	0	>99,9	100	>99,9

Com base nos dados fornecidos por esse estudo de linha genética com nove execuções, o NextSeq 550Dx Instrument pode, consistentemente, sequenciar:

- ▶ Conteúdo GC (Guanine-Cytosine, Guanina-Citosina)  $\geq 19\%$  (todas as bases identificadas em 819 amplicons sequenciados com conteúdo GC de 19% foram corretamente identificadas, com taxa de não identificação de 0,6%)
- ▶ Conteúdo GC  $\leq 87\%$  (todas as bases identificadas em 819 amplicons sequenciados com conteúdo GC de 87% foram corretamente identificadas, com zero não identificação)
- ▶ Comprimentos de PoliA  $\leq 9$  (todas as bases identificadas em 819 amplicons sequenciados contendo uma repetição de PoliA de nove nucleotídeos foram corretamente identificadas, com zero não identificação)
- ▶ Comprimentos de PoliT  $\leq 10$  (todas as bases identificadas em 819 amplicons sequenciados contendo uma repetição de PoliT de dez nucleotídeos foram corretamente identificadas, com zero não identificação)
- ▶ Comprimentos de PoliG  $\leq 7$  (todas as bases identificadas em 819 amplicons sequenciados contendo uma repetição de PoliG de sete nucleotídeos foram corretamente identificadas, com taxa de não identificação de 1,0%)
- ▶ Comprimentos de PoliC  $\leq 6$  (todas as bases identificadas em 2457 amplicons sequenciados contendo uma repetição de PoliC de seis nucleotídeos foram corretamente identificadas, com zero não identificação)
- ▶ Comprimentos de repetição de dinucleotídeos  $\leq 11x$  (todas as bases identificadas em 819 amplicons sequenciados contendo uma repetição de dinucleotídeos 11x foram corretamente identificadas, com taxa de não identificação de 0,5%)
- ▶ Comprimentos de repetição de trinucleotídeos  $\leq 5x$  (todas as bases identificadas em 819 amplicons sequenciados contendo uma repetição de trinucleotídeos 5x foram corretamente identificadas, com taxa de não identificação de 0,5%)
- ▶ Comprimentos de inserção  $\leq 24$  (66343 de 66370 bases identificadas em 819 amplicons sequenciados contendo uma inserção de 24 nucleotídeos foram corretamente identificadas, com taxa de não identificação de 1,2%; nenhuma identificação incorreta ocorreu na região que continha a inserção de 24 nucleotídeos)
- ▶ Comprimentos de deleção  $\leq 25$  (todas as bases identificadas em 2457 amplicons sequenciados contendo uma deleção de 25 nucleotídeos foram corretamente identificadas, com zero não identificação)

### Somática

O estudo descrito aqui foi usado para avaliar a exatidão da identificação de variante do módulo de variante somática no NextSeq 550Dx Instrument com o uso do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

O estudo utilizou um ensaio representativo criado para consultar vários genes, abrangendo 12.588 bases (150 amplicons) em 23 cromossomos diferentes. Foi extraído o DNA Genome Platinum de blocos tratados com FFPE para gerar seis amostras exclusivas para avaliação no estudo.

O DNA da amostra GM12877 foi diluído com o DNA da amostra GM12878 para criar o GM12877-D5 e o GM12877-D7 como um conjunto de variantes heterozigóticas exclusivas com frequências de variante próximas a 5% e 7%. O DNA da amostra GM12878 foi diluído da mesma forma com o DNA da amostra GM12877 para criar o GM12878-D5 e o GM12878-D7. Cada uma das amostras foi testada três vezes, exceto as amostras

diluídas, que foram testadas em réplicas de seis. Foram conduzidas nove execuções usando três instrumentos de sequenciamento, três lotes de reagentes e três operadores ao longo de cinco dias de execução. Foi determinada a exatidão para SNVs, inserções e deleções pela comparação dos resultados com um método de referência composto bem caracterizado, Platinum Genomes versão 2016-1.0. Foram definidas regiões genômicas de confiança com base nesse método de referência, a menos que especificado de outra forma.

Tabela 13 Resumo da concordância somática

Critérios	Total de observações <sup>1</sup>	Resultado por observação <sup>2</sup>	Resultado por execução <sup>3</sup>
PPA para SNV	378	98,9	99,9
PPA para inserções	378	96,9	99,9
PPA para deleções	378	97,1	99,9
NPA	378	>99,9	>99,9
OPA	378	>99,9	>99,9

<sup>1</sup>Calculado como o número de amostras por execução (42) x número de execuções (9) = 378.

<sup>2</sup>Valor mais baixo observado por réplica de amostra em todas as 9 execuções.

<sup>3</sup>Valor mais baixo quando os dados de cada execução são analisados em conjunto.

A **Tabela 14** contém os dados do estudo apresentados com concordância percentual positiva e negativa para cada amostra, sendo que os resultados da variante são comparados com o método de referência composto bem caracterizado para cálculos de PPA. Os três tipos de variante (SNVs, inserções e deleções) são combinados. Como o método de referência só fornece resultados para as variantes de nucleotídeo único e inserções/deleções, os resultados de base não variantes são comparados com sequências de referências do genoma humano versão hg19 para os cálculos do NPA.

Tabela 14 Concordância somática por amostra

Amostra	Taxa de identificação média	Provável	TP	FN	Não identificações de variante	TN	FP	PPA	NPA	OPA
GM12877	98,7	2052	2025	0	27	318682	15	100	>99,9	>99,9
GM12878	98,8	3645	3564	0	81	317645	0	100	100	100
GM12879	99,8	2592	2538	0	54	323614	2	100	>99,9	>99,9
GM12884	99,8	3078	3024	0	54	322038	5	100	>99,9	>99,9
GM12885	99,8	3294	3213	0	81	322121	0	100	100	100
GM12888	99,8	2916	2889	0	27	323048	2	100	>99,9	>99,9
GM12877-D5	99,8	9288	8930	0	358	630621	0	100	100	100
GM12877-D7	99,7	9288	9032	0	256	629719	0	100	100	100
GM12878-D5	99,5	9288	8699	42	547	628582	0	99,5	100	>99,9
GM12878-D7	99,7	9288	9108	0	180	629803	0	100	100	100

A **Tabela 15** contém os dados do estudo apresentados para cada amostra, sendo que os resultados da variante são comparados com o método de referência composto bem caracterizado. A detecção é avaliada para cada tipo de variante: SNVs, inserções e deleções, separadamente. As posições de referência são excluídas.

Tabela 15 Concordância somática por amostra por tipo de variante

Amostra	SNVs			Inserções			Deleções		
	Provável	TP	FN	Provável	TP	FN	Provável	TP	FN
GM12877	999	999	0	567	567	0	486	459	0
GM12878	2457	2457	0	540	513	0	648	594	0

Amostra	SNVs			Inserções			Deleções		
	Provável	TP	FN	Provável	TP	FN	Provável	TP	FN
GM12879	1539	1539	0	567	540	0	486	459	0
GM12884	1836	1836	0	675	648	0	567	540	0
GM12885	2025	2025	0	675	648	0	594	540	0
GM12888	1782	1782	0	621	621	0	513	486	0
GM12877-D5	5454	5392	0	1782	1647	0	2052	1891	0
GM12877-D7	5454	5406	0	1782	1728	0	2052	1898	0
GM12878-D5	5454	5192	28	1782	1651	9	2052	1856	5
GM12878-D7	5454	5445	0	1782	1719	0	2052	1944	0

As dez amostras foram analisadas novamente para a identificação de pequenas inserções e deleções (indels) (Tabela 16). Foi encontrado um total de 71 indels, com tamanho variando de 1–24 bp para inserções e 1–25 bp para deleções.

Tabela 16 Resumo da detecção de indels somáticos

Tipo de variante	Variante prováveis	TP	FN	Não identificações de variante	PPA
Inserção	10773	10282	9	482	99,2
Deleção	11502	10667	5	830	>99,9

Os 150 amplicons se destinaram a cobrir vários conteúdos genômicos. O conteúdo GC dos amplicons variou de 0,19–0,87%. Os amplicons também tiveram uma faixa de nucleotídeos únicos (p. ex., PoliA, PoliT), dinucleotídeos e repetições de trinucleotídeos. Os dados foram comparados para cada amplicon (Tabela 17) para determinar o efeito do conteúdo genômico no percentual de identificações corretas. O percentual de identificações corretas consiste em identificações de variância e de referência e é inferior a 100% no caso de identificações incorretas ou não identificações.

Tabela 17 Exatidão no nível de amplicons somáticos

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
1	1	36450499	36450591	93	93	Indel	0,22	35066	0	88	99,7
2	1	109465122	109465200	79	79	PoliA (5), PoliC (5), indel	0,38	29827	0	35	99,9
3	1	218353867	218353957	91	91	Indel	0,4	34202	0	283	99,2
4	1	223906657	223906748	92	92	Indel	0,49	34613	0	163	99,5
5	1	228526602	228526682	81	81	PoliG (5)	0,69	30571	0	47	99,8
6	1	236372039	236372108	70	70	PoliT (10), indel	0,39	26452	0	8	100,0
7	1	247812041	247812128	88	88	PoliA (5), CT (3), TAA(3), indel	0,27	33148	0	116	99,7
8	2	55862774	55862863	90	90	Indel	0,28	33928	0	92	99,7
9	2	87003930	87004009	80	80	Indel	0,38	30218	0	22	99,9
10	2	177016721	177016805	85	81	N/A	0,65	30616	0	2	>99,9
11	2	186625727	186625801	75	75	PoliA (8)	0,35	28017	0	499	98,3
12	2	190323504	190323591	88	88	PoliT (5)	0,42	33207	0	57	99,8
13	2	200796740	200796826	87	87	PoliT (5), indel	0,31	32524	9	718	97,8
14	2	212245049	212245139	91	91	PoliT (5), PoliA (6), indel	0,3	33972	0	456	98,7
15	2	228147052	228147144	93	93	N/A	0,43	35051	0	103	99,7
16	2	235016350	235016422	73	73	PoliT (5), indel	0,42	27459	0	136	99,5
17	3	4466229	4466321	93	93	AT(3), indel	0,27	34534	0	620	98,2
18	3	46620561	46620643	83	83	N/A	0,43	31339	0	44	99,9

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
19	3	49851331	49851400	70	70	CT(3), indel	0,49	26373	0	87	99,7
20	3	189713161	189713248	88	88	PoliA (5), PoliT (5), PoliA (9), TG(3)	0,41	32829	0	857	97,5
21	3	190106030	190106104	75	74	Indel	0,57	27925	0	47	99,8
22	4	2233667	2233744	78	78	PoliA (6)	0,26	29327	4	162	99,4
23	4	7780541	7780637	97	97	PoliG (6), PoliT (5), PoliA (5)	0,42	36585	0	117	99,7
24	4	15688604	15688681	78	78	N/A	0,29	29427	0	57	99,8
25	4	56236521	56236586	66	62	PoliA (5), indel	0,36	23356	5	75	99,7
26	4	102839244	102839314	71	69	PoliA (5)	0,46	25942	0	140	99,5
27	4	164446743	164446804	62	62	PoliA (7), indel	0,27	22944	0	560	97,6
28	5	1882081	1882158	78	75	N/A	0,78	28299	0	53	99,8
29	5	14769061	14769144	84	84	GT(3), CCA(3)	0,62	31658	0	94	99,7
30	5	41069808	41069871	64	64	N/A	0,39	24120	0	72	99,7
31	5	74077114	74077196	83	83	PoliA (6), indel	0,3	31297	0	77	99,8
32	5	147475343	147475409	67	67	PoliT (5)	0,37	25277	0	55	99,8
33	5	149323731	149323821	91	91	CT(4), AG(3)	0,55	34308	0	90	99,7
34	5	155662213	155662287	75	75	Indel	0,43	28266	0	163	99,4
35	6	6318713	6318814	102	102	PoliG (6)	0,68	38489	0	67	99,8
36	6	24949983	24950074	92	92	Indel	0,63	34730	0	46	99,9
37	6	31084900	31084999	100	94	GCT(5), indel	0,61	35057	0	483	98,6
38	6	32147987	32148084	98	98	PoliT (5), TCT (3), CTT(3)	0,55	36647	0	406	98,9
39	6	32986864	32986958	95	95	Indel	0,53	35681	0	238	99,3
40	6	33408498	33408583	86	86	PoliC (6)	0,7	32438	0	70	99,8
41	6	41647401	41647495	95	94	PoliG (5), indel	0,61	35441	0	91	99,7

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
42	6	112435865	112435955	91	91	PoliA (5)	0,44	34354	0	44	99,9
43	7	22202076	22202148	73	73	N/A	0,44	27575	0	28	99,9
44	7	66276100	66276187	88	88	Indel	0,35	33060	0	213	99,4
45	7	77365735	77365821	87	87	PoliA (7), AG (4)	0,26	32423	0	489	98,5
46	7	110939946	110940030	85	85	Indel	0,38	32074	0	56	99,8
47	7	128533468	128533557	90	90	PoliG (5), indel	0,62	33791	0	281	99,2
48	7	149503875	149503965	91	91	PoliG (6), PoliC (6), indel	0,71	34316	0	82	99,8
49	7	154404519	154404599	81	66	N/A	0,31	24901	0	47	99,8
50	7	156476507	156476599	93	93	Indel	0,35	35067	0	87	99,8
51	8	1817312	1817394	83	83	N/A	0,42	31365	0	9	>99,9
52	8	24811020	24811109	90	89	PoliG (7), CTC (4), indel	0,61	32781	0	890	97,4
53	8	76518625	76518691	67	67	Indel	0,3	25228	0	146	99,4
54	9	103054909	103055006	98	98	PoliG (6)	0,67	36968	0	76	99,8
55	9	105586150	105586214	65	65	Indel	0,32	24472	0	100	99,6
56	9	107620823	107620918	96	96	N/A	0,49	36203	0	85	99,8
57	9	123769149	123769231	83	83	AT(3)	0,37	31329	0	45	99,9
58	9	138995345	138995441	97	97	PoliC (6), indel	0,68	36472	0	201	99,5
59	10	5987120	5987198	79	78	PoliG (5), indel	0,47	29473	0	11	>99,9
60	10	11784629	11784726	98	91	GC(3)	0,87	34188	0	213	99,4
61	10	27317777	27317855	79	79	PoliT (5)	0,3	29843	0	19	99,9
62	10	33018351	33018440	90	90	PoliA (5), PoliT (5)	0,2	33968	0	68	99,8
63	10	45084159	45084253	95	95	Indel	0,35	35829	0	81	99,8
64	10	55892599	55892687	89	88	AC(11), indel	0,42	32098	88	2048	93,8

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
65	10	101611250	101611329	80	80	N/A	0,49	30217	0	28	99,9
66	10	118351373	118351453	81	81	N/A	0,51	30531	0	96	99,7
67	11	8159816	8159912	97	96	N/A	0,45	36105	0	192	99,5
68	11	30177648	30177717	70	70	Indel	0,46	26318	0	153	99,4
69	11	47470345	47470444	100	100	N/A	0,65	37785	0	24	99,9
70	11	59837679	59837740	62	62	Indel	0,37	23368	0	68	99,7
71	11	64418856	64418957	102	102	N/A	0,59	38546	0	10	>99,9
72	11	93529612	93529684	73	73	PoliA (5)	0,4	27516	0	78	99,7
73	11	101347052	101347136	85	85	N/A	0,42	32083	0	48	99,9
74	11	102477336	102477426	91	91	PoliG (6)	0,55	34047	0	369	98,9
75	11	118406285	118406369	85	85	Indel	0,53	32065	0	74	99,8
76	11	120357801	120357885	85	85	PoliA (5), CA (3), indel	0,34	32083	0	47	99,9
77	11	125769313	125769397	85	85	GA(3)	0,52	32103	0	27	99,9
78	12	2834770	2834853	84	84	PoliC (5), indel	0,52	31645	16	525	98,3
79	12	26811004	26811096	93	93	PoliA (7), AC (4)	0,33	34824	0	330	99,1
80	12	30881766	30881846	81	81	N/A	0,49	30497	0	121	99,6
81	12	88474105	88474175	71	71	PoliA (6)	0,35	26773	0	65	99,8
82	12	120966872	120966966	95	95	PoliG (5)	0,68	35830	9	72	99,8
83	13	24167504	24167576	73	73	N/A	0,52	27498	0	114	99,6
84	13	25816961	25817049	89	88	PoliA (5), PoliT (7), PoliA (7), indel	0,22	32824	0	566	98,3
85	13	44880112	44880200	89	89	Indel	0,49	33574	0	77	99,8
86	13	77665218	77665294	77	77	Indel	0,39	29075	0	31	99,9
87	14	31619327	31619393	67	67	GA(3),TA(3)	0,39	25313	0	13	99,9



Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
88	14	39517884	39517966	83	83	N/A	0,25	31360	0	22	99,9
89	14	46958962	46959034	73	72	PoliT (5), indel	0,19	26499	0	717	97,4
90	14	58050030	58050110	81	81	Indel	0,38	30494	0	133	99,6
91	14	82390559	82390649	91	91	Indel	0,35	34313	0	86	99,7
92	14	92549544	92549609	66	66	PoliA (5)	0,41	24555	0	1527	94,1
93	14	102808496	102808589	94	94	Indel	0,62	35472	0	69	99,8
94	15	43170751	43170848	98	96	PoliC (5)	0,45	36264	0	24	99,9
95	15	63446149	63446216	68	68	Indel	0,25	25667	0	37	99,9
96	15	77879807	77879901	95	93	PoliG (5), indel	0,68	34745	0	432	98,8
97	15	81625334	81625428	95	95	PoliT (6)	0,43	35870	0	40	99,9
98	15	85438263	85438334	72	71	Indel	0,65	26762	0	76	99,7
99	15	89817413	89817503	91	91	N/A	0,36	34286	0	112	99,7
100	15	89864274	89864343	70	70	Indel	0,56	26449	0	11	>99,9
101	16	1894910	1894972	63	63	N/A	0,27	23809	0	5	>99,9
102	16	28997904	28997998	95	95	PoliC (5)	0,67	35860	0	50	99,9
103	16	53682908	53682994	87	87	TA(3)	0,41	32835	0	60	99,8
104	16	57954406	57954509	104	104	PoliC (5)	0,67	39177	0	144	99,6
105	16	85706375	85706465	91	91	PoliT (5), indel	0,37	34075	0	323	99,1
106	17	3563920	3564008	89	89	GC(3)	0,64	33632	0	11	>99,9
107	17	3594191	3594277	87	87	PoliC (5), indel	0,67	32752	0	134	99,6
108	17	3970090	3970180	91	91	Indel	0,46	34343	0	82	99,8
109	17	16084945	16085037	93	93	Indel	0,26	35077	0	78	99,8
110	17	33998759	33998849	91	89	PoliT (5)	0,54	33553	0	89	99,7
111	17	39589691	39589774	84	82	PoliA (13), indel (x2)	0,29	30554	53	2296	92,9
112	17	41244394	41244484	91	91	PoliA (5)	0,34	34360	0	38	99,9

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
113	17	45438866	45438957	92	92	PoliA (7), AT (3), AT(4), AT (4), indel	0,26	34367	0	418	98,8
114	17	61502432	61502510	79	79	Indel	0,41	29751	0	119	99,6
115	17	64023582	64023667	86	86	PoliT (7)	0,22	32176	0	340	99,0
116	17	72308237	72308320	84	84	GAG(3)	0,62	31604	7	141	99,5
117	18	2616456	2616522	67	67	GA(3)	0,31	25273	8	45	99,8
118	18	6980478	6980568	91	91	N/A	0,37	34386	0	12	>99,9
119	18	9888026	9888094	69	69	PoliA (6), TG(3)	0,43	25692	0	399	98,5
120	18	38836999	38837073	75	75	PoliA (5), indel	0,37	27923	0	893	96,9
121	18	47405382	47405462	81	81	CTC(3), indel	0,47	30598	0	20	99,9
122	18	54815665	54815749	85	85	CT(3), indel	0,45	31969	0	161	99,5
123	18	59773996	59774060	65	65	N/A	0,48	24531	0	48	99,8
124	19	625143	625241	99	99	N/A	0,59	37298	0	124	99,7
125	19	18121418	18121491	74	74	N/A	0,68	27881	0	109	99,6
126	19	18186574	18186643	70	70	N/A	0,64	26442	0	26	99,9
127	20	746056	746149	94	94	N/A	0,61	35501	0	31	99,9
128	20	10633195	10633276	82	82	AC(3)	0,59	30951	0	72	99,8
129	20	17705633	17705708	76	76	CT(3)	0,58	28686	0	42	99,9
130	20	21766821	21766890	70	70	GT(3),TG(4), indel	0,46	26372	0	88	99,7
131	20	25278421	25278521	101	101	Indel	0,63	38159	0	20	99,9
132	20	50897302	50897368	67	67	Indel	0,36	25188	0	544	97,9
133	20	62331904	62331994	91	88	PoliG (6)	0,73	32969	0	309	99,1
134	20	62690860	62690946	87	87	Indel	0,57	32818	0	77	99,8
135	21	30300823	30300888	66	66	Indel	0,35	24758	9	181	99,2
136	21	33694176	33694273	98	98	PoliT (6), CA(3)	0,54	36902	0	160	99,6

Amplicon	Cromosomo	Início do amplicon	Final do amplicon	Tamanho do fragmento analisado	Bases em regiões de confiança	Conteúdo genômico do amplicon	Conteúdo GC	Identificações corretas	Identificações incorretas	Não identificações	% de identificações corretas
137	21	36710706	36710792	87	87	GT(3), indel	0,39	32841	0	48	99,9
138	21	46644924	46644992	69	69	PoliA (6), AG (3), indel	0,32	25939	0	280	98,9
139	21	46705575	46705664	90	90	PoliT (5), PoliA (6)	0,5	33942	0	78	99,8
140	22	25750774	25750873	100	100	Indel	0,63	37733	0	86	99,8
141	22	32439233	32439329	97	97	N/A	0,68	36617	0	49	99,9
142	22	37409844	37409940	97	97	Indel	0,46	36525	0	162	99,6
143	22	37637596	37637694	99	99	N/A	0,6	37398	0	24	99,9
144	22	47081347	47081438	92	92	Indel	0,66	34754	0	22	99,9
145	X	15870424	15870492	69	69	PoliT (5)	0,26	26046	0	36	99,9
146	X	135288543	135288611	69	69	PoliC (5)	0,62	26019	0	63	99,8
147	X	135290777	135290847	71	71	N/A	0,52	26780	0	58	99,8
148	Y	2655397	2655461	65	0	N/A	0,55	0	0	0	NA
149	Y	2655519	2655609	91	0	N/A	0,48	0	0	0	NA
150	Y	2655609	2655679	71	0	PoliA (5)	0,37	0	0	0	NA

Os resultados do sequenciamento da amostra GM12878 foram comparados com um genótipo de alta confiança para NA12878, estabelecido pelo National Institutes of Standards and Technology (NIST) (v.2.19). Entre os 150 amplicons, 92 estavam totalmente contidos nas regiões genômicas de alta confiança, 41 amplicons tinham sobreposição parcial e 17 amplicons não tinham sobreposição na sequência NIST. Esse resultado acarretou 10.000 coordenadas por réplica para comparação. Identificações de base não variantes foram comparadas com sequências de referências do genoma humano versão hg19. Os resultados da exatidão são mostrados na **Tabela 18**.

Tabela 18 Concordância somática da amostra GM12878 com o banco de dados do NIST

Amostra	N.º de amplicons	Taxa de identificação média	TP	FN	TN	FP	PPA	NPA	OPA
GM12878	133	98,8	2808	0	258488	0	100	100	100

Com base nos dados fornecidos por esse estudo somático com nove execuções, o NextSeq 550Dx Instrument pode, consistentemente, sequenciar:

- ▶ Conteúdo GC  $\geq 19\%$  (todas as bases identificadas em 378 amplicons sequenciados com conteúdo GC de 19% foram corretamente identificadas, com taxa de não identificação de 2,6%)
- ▶ Conteúdo GC  $\leq 87\%$  (todas as bases identificadas em 378 amplicons sequenciados com conteúdo GC de 87% foram corretamente identificadas, com taxa de não identificação de 0,6%)
- ▶ Comprimentos PoliA  $\leq 9$  (todas as bases identificadas em 378 amplicons sequenciados contendo uma repetição de PoliA de nove nucleotídeos foram corretamente identificadas, com taxa de não identificação de 2,5%)
- ▶ Comprimentos PoliT  $\leq 10$  (todas as bases identificadas em 378 amplicons sequenciados contendo uma repetição de PoliT de dez nucleotídeos foram corretamente identificadas, com taxa de não identificação de 0,1%)
- ▶ Comprimentos PoliG  $\leq 6$  (todas as bases identificadas em 2268 amplicons sequenciados contendo uma repetição de PoliG de seis nucleotídeos foram corretamente identificadas, com taxa de não identificação de 0,5%)
- ▶ Comprimentos PoliC  $\leq 6$  (todas as bases identificadas em 756 amplicons sequenciados contendo uma repetição de PoliC de seis nucleotídeos foram corretamente identificadas, com taxa de não identificação de 0,4%)
- ▶ Comprimentos de repetição de dinucleotídeos  $\leq 4x$  (todas as bases identificadas em 1890 amplicons sequenciados contendo uma repetição de dinucleotídeos 4x foram corretamente identificadas, com taxa de não identificação de 0,9%)
- ▶ Comprimentos de repetição de trinucleotídeos  $\leq 5x$  (todas as bases identificadas em 378 amplicons sequenciados contendo uma repetição de trinucleotídeos 5x foram corretamente identificadas, com taxa de não identificação de 1,4%)
- ▶ Comprimentos de inserção  $\leq 23$  (todas as bases identificadas em 378 amplicons sequenciados contendo uma inserção de 23 nucleotídeos foram corretamente identificadas, com taxa de não identificação de 0,8%)
- ▶ Comprimentos de deleção  $\leq 25$  (todas as bases identificadas em 1134 amplicons sequenciados contendo uma deleção de 25 nucleotídeos foram corretamente identificadas, com taxa de não identificação de 0,7%)

## Precisão

A precisão do NextSeq 550Dx Instrument foi determinada por meio do teste de 13 amostras exclusivas de Platinum Genome com o uso de três instrumentos, três lotes de reagentes e três operadores para gerar nove execuções do sequenciamento ao longo de cinco dias de início. O ensaio representativo, as amostras e o método de referência são os mesmos descritos no estudo de exatidão da linha genética. As contribuições da precisão foram determinadas por uma análise de componentes de variância usando a VAF como resposta variável e calculando os desvios padrão no nível do componente para o instrumento, o lote do reagente, o operador e o dia de início (**Tabela 19**). O número total de observações usadas na análise para cada componente do instrumento, operador ou variabilidade do lote de reagente foi 699, 176 e 235 para SNVs, inserções e deleções, respectivamente.

Tabela 19 Resultados da precisão do NextSeq 550Dx Instrument (desvio padrão)

Componente	Tipo de variante	Desvio padrão (DP, Standard Deviation) do componente		DP total	
		Máx.	Mediana	Máx.	Mediana
Lote	SNV	0,0076	0,0002	0,0833	0,0154
	Inserção	0,0104	0,0000	0,0410	0,0157
	Deleção	0,0046	0,0005	0,0560	0,0187
Instrumento	SNV	0,0114	0,0003	0,0840	0,0153
	Inserção	0,0138	0,0009	0,0407	0,0161
	Deleção	0,0079	0,0008	0,0549	0,0187
Operador	SNV	0,0226	0,0008	0,0841	0,0155
	Inserção	0,0344	0,0010	0,0417	0,0164
	Deleção	0,0083	0,0013	0,0547	0,0187
Dia	SNV	0,0277	0,0012	0,0825	0,0160
	Inserção	0,0235	0,0012	0,0409	0,0169
	Deleção	0,0271	0,0014	0,0548	0,0188

### Comparação de métodos (plataforma de sequenciamento)

Amostras de sangue total e FFPE foram avaliadas no NextSeq 550Dx Instrument e MiSeqDx Instrument usando os fluxos de trabalho da linha genética e somáticos do TruSeq Custom Amplicon Kit Dx. A concordância da frequência da variante para amostras de sangue e FFPE foi avaliada usando vários ensaios representativos.

A [Figura 2](#) mostra a correlação da VAF entre os dois instrumentos para um ensaio representativo e a [Tabela 20](#) resume essa correlação para cada painel de ensaio. Com base na forte correlação entre o MiSeqDx Instrument e o NextSeq 550Dx Instrument, as características de desempenho relacionadas aos fatores pré-analíticos (p. ex., métodos de extração ou substâncias interferentes) são determinadas como sendo aplicáveis em ambos os instrumentos. Consulte o folheto informativo do TruSeq Custom Amplicon Kit Dx para conhecer mais detalhes.

Figura 2 Correlação da VAF entre os instrumentos MiSeqDx e NextSeq 550Dx quanto à FFPE (à esquerda) e amostras de sangue (à direita) usando o ensaio 1

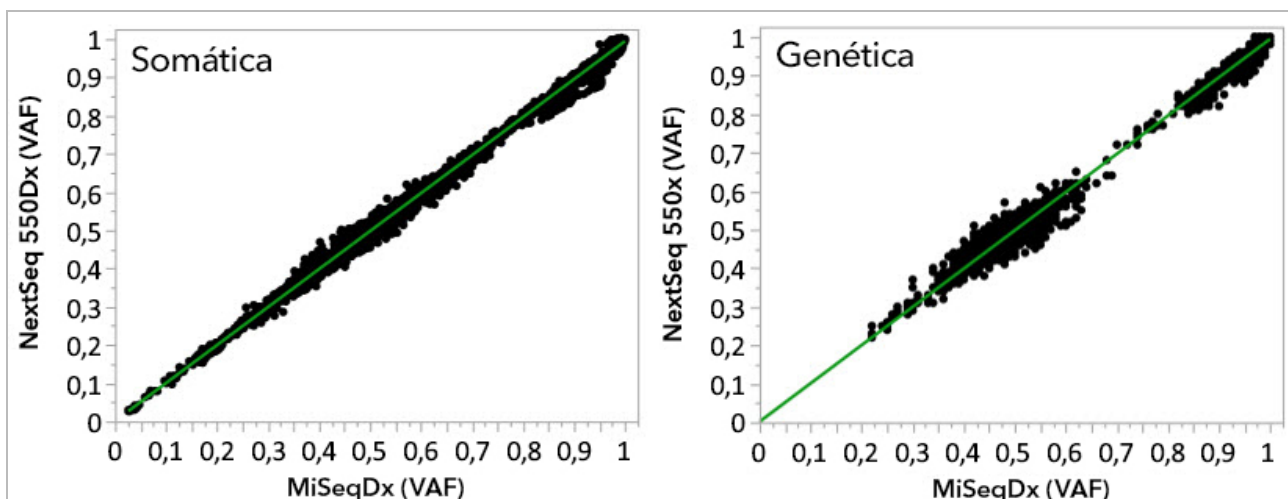


Tabela 20 Resultados da comparação dos métodos usando amostras exclusivas de sangue e FFPE

Origem do gDNA	Ensaio (painel de oligonucleotídeos)	Réplicas biológicas (amostras)	Réplicas técnicas (por amostra)	Observações (n.º de variantes)	Inclinação	Interseção	Correlação (R <sup>2</sup> )
Sangue	Ensaio 1	45	2	8369 <sup>1</sup>	0,992	0,002	0,995 <sup>2</sup>
Sangue	Ensaio 2	45	2	5457	0,995	0,005	0,981
FFPE	Ensaio 1	46	2	8319	0,993	0,000	0,997 <sup>2</sup>
FFPE	Ensaio 3	40	1	280	0,969	0,015	0,978

<sup>1</sup>Dois pontos de dados foram removidos com base na limitação declarada do módulo de variante de linha genética.

<sup>2</sup>Coefficiente de determinação dos gráficos de VAF, conforme ilustrado na Figura 2.

## Reprodutibilidade

A reprodutibilidade do NextSeq 550Dx Instrument foi avaliada com o uso de amostras de Platinum Genome com um ensaio representativo criado para consultar uma variedade de genes, cobrindo 12.588 bases em 23 cromossomos diferentes usando 150 amplicons. Os testes da linha genética consistiram em sete réplicas de 13 amostras; os testes somáticos consistiram em seis réplicas de sete amostras em níveis diferentes de VAF. As amostras foram preparadas com o uso do TruSeq Custom Amplicon Kit Dx.

Os testes foram realizados em três locais externos com o uso de um lote do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Um único NextSeq 550Dx Instrument foi usado em cada local. Dois operadores realizaram os testes em cada local. Cada operador executou os testes em três dias iniciais não consecutivos para cada tipo de amostra para um total de 36 execuções nos três locais. Esses testes resultaram em 18 execuções para cada um dos fluxos de trabalho somáticos e da linha genética.

### Linha genética

Variantes de linha genética com nível de VAF  $\geq 0,2$  são relatadas como positivas (variantes). Para variantes de linha genética supostamente positivas, os dados foram avaliados quanto à taxa de não identificação e à taxa de identificação positiva correta em cada tipo de variante (SNV, inserção, deleção). A [Tabela 21](#) resume os dados observados, juntamente com os níveis de confiança de 95% superiores e inferiores (LCL/UCL) calculados pelo método da pontuação de Wilson para cada tipo de variante.

Tabela 21 Observações da identificação da linha genética para resultados positivos prováveis por tipo de variante

Tipo de variante	Não identificação			Identificação positiva correta				
	Observada	Total	Percentual	Observada	Total	Percentual	LCL 95%	UCL 95%
SNV	16	110.376	0,014	110.349	110.360	99,99	99,98	99,99
Inserções	1026	37.044	2,77	36.018	36.018	100	99,99	100,00
Deleções	648	34.776	1,86	34.128	34.128	100	99,99	100,00

Variantes de linha genética com nível de VAF  $< 0,2$  são relatadas como negativas (do tipo selvagem). Para locais de linha genética supostamente negativos, os dados foram avaliados quanto à não identificação e às taxas de identificação do tipo selvagem corretas. A [Tabela 22](#) resume os dados observados, juntamente com os níveis de confiança de 95% superiores e inferiores (LCL/UCL) calculados pelo método da pontuação de Wilson.

Tabela 22 Observações da identificação da linha genética para resultados negativos prováveis

Tipo de variante	Não identificação			Identificação negativa correta				
	Observada	Total	Percentual	Observada	Total	Percentual	LCL 95%	UCL 95%
Tipo selvagem	4883	19.600.182	0,025	19.595.299	19.595.299	100	100,00	100,00

As variantes de linha genética com nível de VAF  $\geq 0,2$  e  $< 0,7$  são identificadas como heterozigóticas positivas para a variante e as variantes com nível de VAF  $\geq 0,7$  são identificadas como homozigóticas positivas para a variante. Foram usadas amostras com variantes heterozigóticas para determinar se a variabilidade inerente do ensaio afetaria a identificação do genótipo. O Cx foi determinado para ambos os cortes (0,2 para genótipos heterozigóticos e 0,7 para genótipos homozigóticos), sendo x a proporção de testes repetidos que excedem o corte. Para o corte inferior com VAF de 0,2, o Cx foi  $\geq 99,999\%$ , indicando que  $\geq 99,999\%$  das variantes heterozigóticas seriam identificadas como heterozigóticas. Com relação ao corte superior com VAF de 0,7, o Cx foi  $\leq 0,001\%$ , indicando, portanto, que  $\leq 0,001\%$  das variantes heterozigóticas seriam identificadas como homozigóticas. A **Tabela 23** resume os resultados por tipo de variante.

As variantes de linha genética com nível de VAF  $\geq 0,2$  e  $< 0,7$  são identificadas como heterozigóticas positivas para a variante e as variantes com nível de VAF  $\geq 0,7$  são identificadas como homozigóticas positivas para a variante. Foram usadas amostras com variantes heterozigóticas para determinar se a variabilidade inerente do ensaio afetaria a identificação do genótipo. O Cx foi determinado para ambos os cortes (0,2 para genótipos heterozigóticos e 0,7 para genótipos homozigóticos), sendo x a proporção de testes repetidos que excedem o corte. Com relação ao corte inferior com VAF de 0,2, o Cx foi  $\geq 99,999\%$ , indicando, portanto, que  $\geq 99,999\%$  das variantes heterozigóticas seriam identificadas como heterozigóticas. Para o corte superior com VAF de 0,7, o Cx foi  $\leq 0,001\%$ , indicando que  $\leq 0,001\%$  das variantes heterozigóticas seriam identificadas como homozigóticas. A **Tabela 23** resume os resultados por tipo de variante.

Tabela 23 Valores de Cx da linha genética para variantes heterozigóticas

Tipo de variante	VAF de corte de 0,2	VAF de corte de 0,7
	$\geq C99,999\%$	$\leq C0,001\%$
SNV	94/94	94/94
Inserções	24/24	24/24
Deleções	35/35	35/35
Total	153	153

### Somática

Variantes somáticas com nível de VAF  $\geq 0,026$  são relatadas como positivas (variantes). As observações com níveis de VAF  $\geq 0,01$  e  $< 0,026$  foram consideradas ambíguas para esta análise (nem positivas nem negativas, marcadas como baixa frequência de variante). Para avaliar o desempenho, os resultados foram calculados de três formas:

- ▶ Melhor caso: todos os resultados ambíguos foram considerados como identificação positiva correta (concordância com os resultados prováveis)
- ▶ Pior caso: todos os resultados ambíguos foram considerados como identificação incorreta (discordância com os resultados prováveis)
- ▶ Caso de exclusão: todos os resultados ambíguos foram excluídos da análise

Três tabelas, a **Tabela 24**, a **Tabela 25** e a **Tabela 26**, resumem os resultados de identificação para o melhor caso, o pior caso e o caso de exclusão, respectivamente, junto com os níveis inferior e superior de 95% de confiança (LCL/UCL), calculados pelo método da pontuação de Wilson.

Tabela 24 Observações da identificação somática para resultados positivos prováveis por tipo de variante (melhor caso)

Tipo de variante	Identificação positiva correta				
	Observada	Total	Percentual	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserções	18.036	18.036	100	99,98	100,00
Deleções	18.381	18.381	100	99,98	100,00

Tabela 25 Observações da identificação somática para resultados positivos prováveis por tipo de variante (pior caso)

Tipo de variante	Identificação positiva correta				
	Observada	Total	Percentual	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserções	18.000	18.036	99,8	99,72	99,86
Deleções	18.381	18.381	100	99,98	100,00

Tabela 26 Observações da identificação somática para resultados positivos prováveis por tipo de variante (identificações ambíguas removidas)

Tipo de variante	Identificação positiva correta				
	Observada	Total	Percentual	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserções	18.000	18.000	100	99,98	100,00
Deleções	18.381	18.381	100	99,98	100,00

Variantes somáticas com nível de VAF <0,01 são relatadas como identificação negativa (do tipo selvagem). Para locais de linha genética supostamente negativos, os dados foram avaliados quanto à taxa de não identificação e à taxa de identificação do tipo selvagem correta. As identificações do tipo selvagem corretas foram determinadas excluindo as não identificações e subtraindo as identificações observadas que se enquadraram na zona ambígua (níveis de VAF  $\geq 0,01$  e  $< 0,026$ ) e também as identificações incorretas que estavam acima do corte (níveis de VAF  $\geq 0,026$ ) do total. A [Tabela 27](#) resume os resultados observados, totais e percentuais para locais somáticos negativos para taxa de não identificação e taxa de identificação do tipo selvagem correta, juntamente com os níveis de confiança de 95% superiores e inferiores (LCL/UCL) calculados pelo método da pontuação de Wilson.

Tabela 27 Observações da identificação somática para resultados negativos prováveis

Tipo de variante	Não identificação			Identificação correta		Correta	Total	Percentual	LCL 95%	UCL 95%
	Observada	Total	Percentual	Ambígua	Incorreta					
Tipo selvagem	36.326	8.909.676	0,408	2254	121	8.870.975	8.873.350	99,97	99,972	99,974

Foram avaliadas amostras somáticas em diferentes níveis de VAF para a mesma variante para determinar o C95 do ensaio (em cada tipo de variante). Para avaliar a variabilidade perto do corte do ensaio, foram usadas amostras que tinham nível provável de VAF entre 0,02 e 0,07. Foi determinado o C95 de cada variante, com o C95 mais alto para cada tipo de variante mostrado na [Tabela 28](#).

Tabela 28 Resumo de C95 somático

Tipo de variante	N	C95
SNV	74	0,0613
Inserção	24	0,0573
Deleção	33	0,0575



## Desempenho do NextSeq 550Dx High Output Reagent Kit v2.5 (300 Cycle)

### Visão geral

O NextSeq 550Dx é usado com dois kits de reagentes: NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Para demonstrar que o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) pode atender aos requisitos de desempenho analítico verificados e validados com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles), foram conduzidos estudos com o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Duas preparações de biblioteca usando o TruSeq Custom Amplicon Kit Dx foram realizadas, uma com o fluxo de trabalho de variante de linha genética e outra com o fluxo de trabalho de variante somática. Foram testadas bibliotecas de cada fluxo de trabalho com três lotes do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) em três instrumentos NextSeq 550Dx. Além disso, os testes de cada fluxo de trabalho incluíram uma execução única com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

### Sensibilidade analítica (limite de branco [LoB] e limite de detecção [LoD])

A verificação com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) demonstrou que o NextSeq 550Dx Instrument foi capaz de detectar variantes em uma VAF de 0,05 com erro do tipo II  $\leq 0,05$  e que a VAF de corte de 0,026 usada pelo módulo de variante somática (LoB eficaz) suporta um erro do tipo I  $\leq 0,01$ . Com base nessas características, é esperado que uma variante em uma VAF de 0,05 seja maior ou igual à VAF de 0,026 95% do tempo e que uma posição do tipo selvagem seja menor do que uma VAF de 0,026 99% do tempo. Para garantir que essas características tenham sido atendidas com o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), foram realizadas medições repetidas no NextSeq 550Dx Instrument com amostras do tipo selvagem (amostras LoB) e amostras contendo variantes em uma VAF de 0,05 (amostras LoD) usando o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). A proporção de identificações acima e abaixo do corte de 0,026 foi comparada com as características estabelecidas com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Os testes incluíram duas amostras LoD, cada uma com um conjunto exclusivo de variantes direcionadas para uma VAF de 0,05 e amostras LoB correspondentes que eram do tipo selvagem para as variantes direcionadas. Para a preparação de bibliotecas, amostras LoD e LoB foram processadas em réplicas de oito e sete, respectivamente, usando o TruSeq Custom Amplicon Kit Dx. As bibliotecas foram inicialmente sequenciadas com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) para identificar variantes/coordenadas genômicas para avaliação de LoB/LoD com o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Todas as variantes com VAF média entre 0,045 e 0,055 (variantes LoD) com base nos resultados do NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) foram usadas para análise de LoD (N = 51 variantes). Para análise de LoB, as 51 coordenadas genômicas correspondentes foram avaliadas.

Para a avaliação do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), as bibliotecas foram sequenciadas em três execuções durante três dias consecutivos usando o mesmo instrumento e o mesmo lote do kit de reagentes. Esses testes totalizaram 24 réplicas para cada uma das 51 variantes LoD e 21 réplicas para cada uma das posições do tipo selvagem correspondentes. A proporção das identificações do tipo selvagem com VAF  $< 0,026$  é fornecida na [Tabela 29](#). A proporção de identificações de variantes LoD com VAF maior ou igual a 0,026 é fornecida na [Tabela 30](#).

Tabela 29 Proporção de identificações  $< 0,026$  para posições do tipo selvagem (avaliação de características do LoB)

Tipo de variante	Posições avaliadas	Total de observações	N.º de medições de VAF $\geq 2,6\%$	Proporção $< 2,6\%$	Proporção 95% Intervalo de confiança
SNV	32	672	0	1	0,994 – 1
Inserção	11	231	0	1	0,984 – 1
Deleção	8	168	0	1	0,978 – 1

Tabela 30 Proporção de identificações  $\geq 0,026$  VAF para Variantes LoD (avaliação de características do LoD)

Tipo de variante	Posições avaliadas	Total de observações	N.º de medições de VAF $< 2,6\%$	N.º de medições de VAF $\geq 2,6\%$	Proporção $\geq 2,6\%$	Proporção 95% Intervalo de confiança
SNV	32	768	1	767	0,999	0,993 – 1
Inserção	11	264	3	261	0,989	0,967 – 0,996
Deleção	8	192	2	190	0,99	0,963 – 0,997

## Exatidão

### Linha genética

O estudo a seguir foi conduzido para avaliar a exatidão da identificação de variantes do módulo de variante de linha genética usando o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Doze amostras exclusivas de Platinum Genome foram testadas por um ensaio representativo. Foi conduzido um total de 11 execuções usando três instrumentos NextSeq 550Dx e três NextSeq 550Dx High Output Reagent Kits v2.5 (300 cycles).

Foi determinada a exatidão para SNVs, inserções e deleções comparando os resultados com um método de referência composto bem caracterizado, Platinum Genomes versão 2016-1.0. Os resultados da exatidão de uma única execução do sequenciamento com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) são fornecidos como referência. Um resumo dos resultados é mostrado na [Tabela 31](#).

Tabela 31 Resumo de concordância da linha genética

Crítérios	Total de observações (v2.5) <sup>1</sup>	Resultado por observação (v2.5) <sup>2</sup>	Resultado por observação (v2) <sup>3</sup>	Resultado por execução (v2.5) <sup>4</sup>	Resultado por execução (v2) <sup>4</sup>
PPA para SNV	1056	98,7	98,7	>99,9	>99,9
PPA para inserções	1056	100	100	100	98,9
PPA para deleções	1056	95,2	95,2	>99,9	100
NPA	1056	100	100	100	100
OPA	1056	>99,9	>99,9	>99,9	>99,9

<sup>1</sup>Calculado como número de amostras por execução x número de execuções (96 amostras por execução x 11 execuções = 1056 observações).

<sup>2</sup>Valor mais baixo observado por réplica de amostra em todas as execuções (com base em 11 execuções no NextSeq 550Dx High Output Reagent Kit v2.5).

<sup>3</sup>Valor mais baixo observado por réplica de amostra em uma execução (96 observações no total).

<sup>4</sup>Valor mais baixo quando os dados de cada execução são analisados em conjunto.

### Somática

O estudo a seguir foi conduzido para avaliar a exatidão da identificação de variante do módulo de variante somática no NextSeq 550Dx Instrument com o uso do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Dez amostras FFPE de Platinum Genome (duas com variantes diluídas em VAF de 0,05) foram testadas por meio de um ensaio representativo. Foi conduzido um total de 11 execuções usando três instrumentos NextSeq 550Dx e três lotes do NextSeq 550Dx High Output Reagent Kits v2.5 (300 cycles).

Foi determinada a exatidão para SNVs, inserções e deleções comparando os resultados com um método de referência composto bem caracterizado, Platinum Genomes versão 2016-1.0. Os resultados da exatidão de uma única execução do sequenciamento com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) são fornecidos como referência. Um resumo dos resultados é mostrado na [Tabela 32](#).

Tabela 32 Resumo da concordância somática

Critérios	Total de observações (v2.5) <sup>1</sup>	Resultado por observação (v2.5) <sup>2</sup>	Resultado por observação (v2) <sup>3</sup>	Resultado por execução (v2.5) <sup>4</sup>	Resultado por execução (v2) <sup>4</sup>
PPA para SNV	528	100	100	100	100
PPA para inserções	528	96,9	96,9	>99,9	>99,9
PPA para deleções	528	100	100	100	100
NPA	528	>99,9	>99,9	>99,9	>99,9
OPA	528	>99,9	>99,9	>99,9	>99,9

<sup>1</sup>Calculado como número de amostras por execução x número de execuções (48 amostras por execução x 11 execuções = 528 observações).

<sup>2</sup>Valor mais baixo observado por réplica de amostra em todas as execuções (com base em 11 execuções no NextSeq 550Dx High Output Reagent Kit v2.5).

<sup>3</sup>Valor mais baixo observado por réplica de amostra em uma execução (96 observações no total).

<sup>4</sup>Valor mais baixo quando os dados de cada execução são analisados em conjunto.

## Precisão

### Linha genética

A precisão do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) com o módulo de variante de linha genética foi avaliada com o uso de amostras de Platinum Genome e um ensaio representativo. Os testes consistiram na preparação de uma única biblioteca usando o TruSeq Custom Amplicon Kit Dx e incluíram 12 amostras processadas, cada uma com oito réplicas. As bibliotecas foram sequenciadas com três lotes do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) e três instrumentos NextSeq 550Dx para um total de nove execuções do sequenciamento.

Foram usadas amostras com variantes heterozigóticas para determinar se a variabilidade inerente do ensaio afetaria a identificação do genótipo (N = 153 variantes heterozigóticas exclusivas). O Cx foi determinado para cortes do módulo da variante de linha genética (0,2 para genótipos heterozigóticos e 0,7 para genótipos homozigóticos), sendo x a proporção de testes repetidos que excedem o corte. Para o corte inferior com VAF de 0,2, a variante com Cx mínimo para o NextSeq 550Dx Reagent Kit v2.5 (300 cycles) foi >99,9%, indicando que >99,9% das variantes heterozigóticas seriam identificadas como heterozigóticas. Para o corte superior com VAF de 0,7, a variante com Cx máximo para o NextSeq 550Dx Reagent Kit v2.5 (300 cycles) foi <1,5%, indicando que ≤1,5% das variantes heterozigóticas seriam identificadas como homozigóticas. A Tabela 33 resume os resultados por tipo de variante. Valores de Cx da execução do sequenciamento simples usando o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) são fornecidos para referência.

Tabela 33 Valores de Cx da linha genética para variantes heterozigóticas

Tipo de variante	N	VAF de corte de 0,2		VAF de corte de 0,7	
		Cx mín. (v2.5) <sup>1</sup>	Cx mín. (v2) <sup>2</sup>	Cx máx. (v2.5) <sup>1</sup>	Cx máx. (v2) <sup>2</sup>
SNV	94	>99,9%	>99,9%	1,5%	1,0%
Inserções	24	100%	100%	0%	<0,1%
Deleções	35	100%	>99,9%	<0,1%	<0,1%

<sup>1</sup>Valores de Cx baseados nas estimativas do desvio padrão total provenientes da análise de componentes da variância.

<sup>2</sup>Valores de Cx baseados nos desvios padrão da amostra.

## Somática

A precisão do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) com o módulo de variante somática foi avaliada com o uso de amostras FFPE de Platinum Genome e um ensaio representativo. Os testes consistiram na preparação de uma única biblioteca usando o TruSeq Custom Amplicon Kit Dx e incluíram duas amostras processadas, cada uma com oito réplicas. As bibliotecas foram sequenciadas usando três lotes do NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) e três instrumentos NextSeq 550Dx para um total de nove execuções do sequenciamento.

Foram usadas variantes somáticas com níveis prováveis de VAF  $\leq 0,10$  (N = 131 variantes exclusivas) para avaliar a variabilidade do instrumento próximo à VAF de corte do módulo de variante somática (variantes somáticas com nível de VAF  $\geq 0,026$  são chamadas de positivas para a variante). Foram determinados valores de C95 para cada uma das variantes somáticas. Os valores de C95 representam a VAF na qual a probabilidade de que eles sejam maiores do que a VAF de corte do módulo de variante somática é de 95%. Os valores mais altos de C95 por variante são mostrados na [Tabela 34](#). Resultados de C95 de uma execução do sequenciamento simples usando o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) são fornecidos para referência.

Tabela 34 Resumo de C95 somático

Tipo de variante	N.º de variantes avaliadas	C95 (v2.5) <sup>1</sup>	C95 (v2) <sup>2</sup>
SNV	74	0,064	0,063
Inserções	24	0,062	0,061
Deleções	33	0,060	0,060

<sup>1</sup>Valores de C95 baseados nas estimativas do desvio padrão total provenientes da análise de componentes da variância.

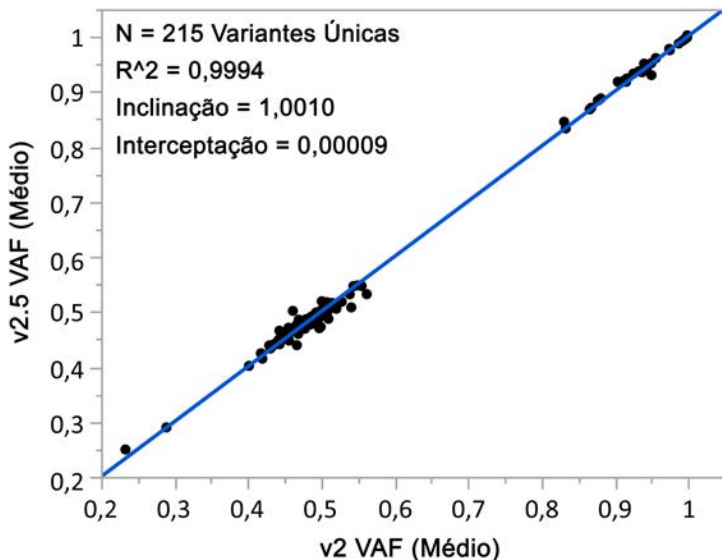
<sup>2</sup>Valores de C95 baseados nos desvios padrão da amostra.

## Comparação de métodos (kit de reagentes)

### Linha genética

As VAFs médias de 215 variantes exclusivas foram avaliadas no NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e no NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) usando resultados gerados no módulo de variante de linha genética. As médias da VAF foram calculadas em 11 execuções do sequenciamento (v2.5) e em uma execução do sequenciamento (v2). Pelo menos oito réplicas foram usadas para calcular a média para cada variante. A [Figura 3](#) mostra a correlação da VAF entre os dois kits de reagentes. Com base na forte correlação linear da VAF e na similaridade dos resultados entre os kits de reagentes, as características de desempenho inicialmente verificadas e validadas com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) com o módulo de variante de linha genética são determinadas como aplicáveis ao NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

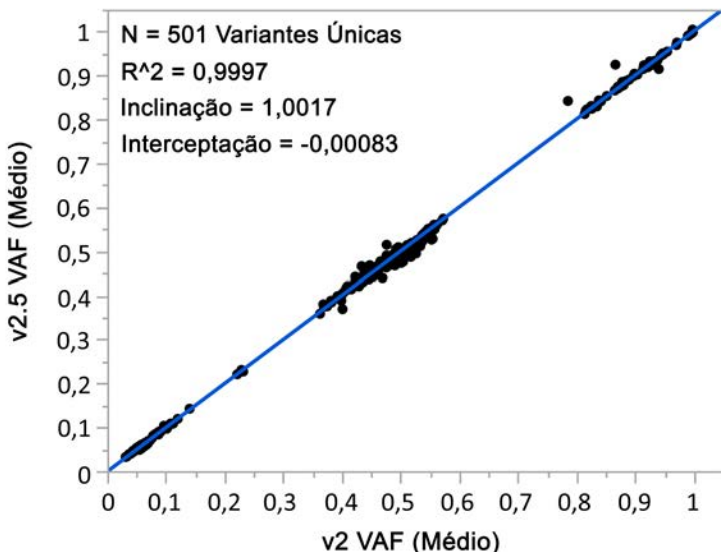
Figura 3 Correlação da frequência alélica da variante (VAF) do módulo de variante de linha genética entre o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).



### Somática

As VAFs médias de 501 variantes exclusivas foram avaliadas no NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e no NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) usando resultados gerados no módulo de variante somática. As médias da VAF foram calculadas em 11 execuções do sequenciamento (v2.5) e em uma execução do sequenciamento (v2). Pelo menos três réplicas foram usadas para calcular a média para cada variante exclusiva. A [Figura 4](#) mostra a correlação da VAF entre os dois kits de reagentes. Com base na correlação da VAF e nos resultados de similaridade entre os kits de reagentes, as características de desempenho verificadas e validadas com o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) com o módulo de variante somática são determinadas como aplicáveis ao NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

Figura 4 Correlação da frequência alélica da variante (VAF) do módulo de variante somática entre o NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e o NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).



## Histórico de revisões

Documento	Data	Descrição da alteração
Documento n.º 1000000030326 v06	Maio de 2022	Foram feitas atualizações para retificar o conteúdo inadvertidamente adicionado do software de origem.
Documento n.º 1000000030326 v05	Novembro de 2021	Adicionada declaração de avisos e precauções sobre o relato de incidentes graves. Adicionada uma declaração a Princípios do procedimento especificando o usuário previsto. Removida referência ao High Output Reagent Kit v2 (300 cycles). Adicionada referência ao High Output Reagent Kit v2.5 (75 cycles).
Documento n.º 1000000030326 v04	Agosto de 2021	Adicionada a tabela Histórico de revisões. Atualizado o endereço do Representante autorizado da UE.

## Patentes e marcas comerciais

Este documento e seu conteúdo são de propriedade da Illumina, Inc. e de suas afiliadas (“Illumina”) e destinam-se exclusivamente ao uso contratual de seu cliente com relação ao uso dos produtos descritos neste documento e para nenhuma outra finalidade. Este documento e seu conteúdo não devem ser usados ou distribuídos para nenhuma outra finalidade nem comunicados, divulgados ou reproduzidos de nenhuma forma sem o consentimento prévio por escrito da Illumina. A Illumina não concede nenhuma licença sob seus direitos de patente, marca registrada, direitos autorais ou lei comum, nem direitos semelhantes de terceiros por meio deste documento.

As instruções neste documento devem ser estrita e explicitamente seguidas por pessoal devidamente treinado e qualificado para garantir o uso adequado e seguro dos produtos descritos neste documento. Todo o conteúdo deste documento deve ser lido e compreendido por completo antes da utilização de tais produtos.

**NÃO LER COMPLETAMENTE E NÃO SEGUIR EXPLICITAMENTE TODAS AS INSTRUÇÕES AQUI CONTIDAS PODE RESULTAR EM DANOS AO(S) PRODUTO(S), FERIMENTOS A PESSOAS, INCLUSIVE USUÁRIOS OU OUTROS, E DANOS A OUTROS BENS, ANULANDO TODA GARANTIA APLICÁVEL AO(S) PRODUTO(S).**

A ILLUMINA NÃO SE RESPONSABILIZA POR QUALQUER PROBLEMA CAUSADO PELO USO INDEVIDO DO(S) PRODUTO(S) MENCIONADO(S) ACIMA (INCLUINDO PARTES SEPARADAS OU O SOFTWARE).

© 2022 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).

## Informações de contato



Illumina  
5200 Illumina Way  
San Diego, California 92122, EUA  
+1 (800) 809-ILMN (4566)  
+1 (858) 202-4566 (fora da América do Norte)  
[techsupport@illumina.com](mailto:techsupport@illumina.com)  
[www.illumina.com](http://www.illumina.com)



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
Países Baixos

**Patrocinador australiano**  
Illumina Australia Pty Ltd  
Nursing Association Building  
Level 3, 535 Elizabeth Street  
Melbourne, VIC 3000  
Austrália

## Rótulos do produto

Para obter uma referência completa dos símbolos que possam ser exibidos na embalagem e na rotulagem do produto, consulte a chave de símbolos do seu kit em [support.illumina.com](http://support.illumina.com).