

# DRAGEN for Illumina DNA Prep with Enrichment Dx on NextSeq 550Dx

Podręcznik użytkownika aplikacji

ZASTRZEŻONE MATERIAŁY FIRMY ILLUMINA

Nr dokumentu: 200025238, wer. 00

Luty 2023 r.

DO STOSOWANIA W DIAGNOSTYCE IN VITRO.

Niniejszy dokument oraz jego treść stanowią własność firmy Illumina, Inc. oraz jej podmiotów zależnych („Illumina”) i są przeznaczone wyłącznie do użytku zgodnego z umową przez klienta firmy w związku z użytkowaniem produktów opisanych w niniejszym dokumencie, z wyłączeniem innych celów. Niniejszy dokument oraz jego treść nie będą wykorzystywane ani rozpowszechniane w innych celach i/lub publikowane w inny sposób, ujawniane ani kopiowane bez pisemnej zgody firmy Illumina. Firma Illumina na podstawie niniejszego dokumentu nie przenosi żadnych licencji podlegających przepisom w zakresie patentów, znaków towarowych czy praw autorskich ani prawu powszechnemu lub prawom pokrewnym osób trzecich.

W celu zapewnienia właściwego i bezpiecznego użytkowania produktów opisanych w niniejszym dokumencie podane instrukcje powinny być ściśle przestrzegane przez wykwalifikowany i właściwie przeszkolony personel. Przed rozpoczęciem użytkowania tych produktów należy zapoznać się z całą treścią niniejszego dokumentu.

**NIEZAPOZNANIE SIĘ LUB NIEDOKŁADNE PRZESTRZEGANIE WSZYSTKICH INSTRUKCJI PODANYCH W NINIEJSZYM DOKUMENCIE MOŻE SPOWODOWAĆ USZKODZENIE PRODUKTÓW LUB OBRAŻENIA CIAŁA UŻYTKOWNIKÓW LUB INNYCH OSÓB ORAZ USZKODZENIE INNEGO MIENIA, A TAKŻE SPOWODUJE UNIEWAŻNIENIE WSZELKICH GWARANCJI DOTYCZĄCYCH PRODUKTÓW.**

**FIRMA ILLUMINA NIE PONOSI ODPOWIEDZIALNOŚCI ZA NIEWŁAŚCIWE UŻYTKOWANIE PRODUKTÓW (W TYM ICH CZĘŚCI I OPROGRAMOWANIA) OPISANYCH W NINIEJSZYM DOKUMENCIE.**

© 2023 Illumina, Inc. Wszelkie prawa zastrzeżone.

Wszystkie znaki towarowe są własnością firmy Illumina, Inc. lub ich odpowiednich właścicieli. Szczegółowe informacje na temat znaków towarowych można znaleźć pod adresem [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

## Historia wersji

<b>Dokument</b>	<b>Data</b>	<b>Opis zmiany</b>
200025238 wer. 00	Luty 2023 r.	Pierwsze wydanie.

## Spis treści

Historia wersji .....	iii
Przegląd .....	1
Metody analizy .....	1
Tworzenie zaplanowanego przebiegu .....	5
Ustawienia .....	7
Plik wykazu .....	8
Filtrowanie szumów (opcjonalnie) .....	9
Dane wyjściowe analizy .....	9
Pliki FASTQ .....	10
Pliki BAM .....	11
Pliki VCF .....	12
Ponowne umieszczanie analizy w kolejce .....	18
<b>Pomoc techniczna .....</b>	<b>20</b>

# Przegląd

Aplikacja DRAGEN for Illumina DNA Prep with Enrichment Dx (DRAGEN for IDPE Dx) służy do planowania i przeprowadzania dodatkowej analizy bibliotek IDPE Dx wygenerowanych do sekwencjonowania w NextSeq 550Dx.

DRAGEN for IDPE Dx obsługuje sekwencjonowanie do analizy w przypadku użycia z przygotowaniem biblioteki Illumina DNA Prep with Enrichment Dx oraz NextSeq 550Dx i Illumina DRAGEN Server for NextSeq 550Dx.

## Metody analizy

DRAGEN for IDPE Dx wykonuje demultipleksowanie, generowanie FASTQ, mapowanie odczytu, dopasowanie do genomu referencyjnego i rozpoznawanie wariantów małych w zależności od wybranej procedury:

- Generowanie pliku FASTQ
- Germline FASTQ i VCF generation
- Somatic FASTQ i VCF generation

**UWAGA** ORA compression jest dostępna do stosowania we wszystkich trzech procedurach. DRAGEN ORA Compression to w pełni bezstratne oprogramowanie kompresujące, które tworzy plik z rozszerzeniem Original Read Archive (\*.ora). Format ora jest referencyjnym formatem kompresji dla plików FASTQ i jest przeznaczony do bardzo szybkiej kompresji/dekompresji i wysokiego współczynnika kompresji.

## Generowanie pliku FASTQ

Zmontowane sekwencje są zapisywane w plikach FASTQ dla każdej próbki. Pliki FASTQ to pliki tekstowe zawierające dane sekwencjonowania i wyniki jakości dla tylko jednej próbki. Dla każdej próbki generowane są oddzielne pliki FASTQ dla pasma komory przepływowej, na odczyt sekwencjonowania. Nazwa próbki określona podczas konfiguracji przebiegu jest zawarta w nazwie pliku FASTQ. Pliki FASTQ to podstawowe dane wejściowe do dopasowywania. Pierwszym etapem generowania FASTQ jest demultipleksowanie. Demultipleksowanie przypisuje klastry, które przejdą przez filtr, do próbki poprzez porównanie każdej sekwencji odczytu indeksu z sekwencjami indeksu określonymi dla danego przebiegu. Na tym etapie nie są brane pod uwagę żadne wartości dotyczące jakości. Odczyty indeksów identyfikuje się poprzez następujące czynności:

- Próbkę są ponumerowane, począwszy od numeru 1, na podstawie kolejności, w jakiej je wymieniono dla przebiegu.
- Próbka numer 0 jest zarezerwowana dla klastrów, które nie zostały przypisane do próbki.

- Klastry są przypisywane do próbki, gdy sekwencja indeksu jest dokładnie zgodna lub gdy na odczyt indeksu przypada maksymalnie jedna niezgodność.

Oprogramowanie zawiera kompresję ORA do kompresowania plików FASTQ. Ten format można opcjonalnie włączyć. W przypadku korzystania z formatu ORA (\*.ora) suma kontrolna md5 zawartości FASTQ jest zachowywana po cyklu kompresji i dekompresji, aby zapewnić kompresję bezstratną.

## Dopasowywanie i mapowanie DNA

Po wygenerowaniu pliku FASTQ odczyty są mapowane i wyrównywane z genomem referencyjnym. Pierwszym etapem mapowania jest wygenerowanie punktów początkowych z odczytu, a następnie wyszukanie dokładnych dopasowań w genomie referencyjnym. Wyniki te są następnie poprawiane poprzez wykonanie pełnych dopasowań Smitha-Watermana w lokalizacjach o najwyższej gęstości dopasowanych punktów początkowych. Ten dobrze udokumentowany algorytm działa poprzez porównanie każdej pozycji odczytu ze wszystkimi pozycjami kandydującymi odniesienia. Porównania te odpowiadają maczyi potencjalnych dopasowań między odczytem a odniesieniem. Dla każdej z tych pozycji kandydujących do dopasowania metoda Smith-Waterman generuje wyniki, które są wykorzystywane do oceny, czy najlepsze dopasowanie przechodzące przez tę komórkę maczyi dociera do niej przez dopasowanie, czy niezgodność nukleotydów (ruch ukośny), delecję (ruch poziomy) lub insercję (ruch pionowy). Dopasowanie między odczytem a odniesieniem zapewnia premię do wyniku, a niedopasowanie lub indel nakłada karę. Ścieżka o najwyższej ogólnej punktacji w maczyi to wybrane dopasowanie. Algorytm jest akcelerowany sprzętowo na kartach DRAGEN programowalnych w terenie macierzy bramek (FPGA). Genom referencyjny używany w aplikacji jest tworzony z UCSC hg19 FASTA z opcją DRAGEN w celu utworzenia tabeli skrótów z mapowaniem alt-aware na podstawie przeniesienia.

## Wykrywanie wariantu linii zarodkowej DRAGEN

DRAGEN Germline Small Variant Caller przyjmuje zmapowane i dopasowane odczyty DNA jako dane wejściowe i rozpoznaje polimorfizmy pojedynczych nukleotydów (SNP) oraz indele poprzez połączenie wykrywania kolumnowego i lokalnego łączenia haplotypów *de novo*. Aby włączyć DRAGEN Germline Small Variant Caller, wybierz procedurę wariantu linii zarodkowej.

Połączenie wariant linii zarodkowej jest zazwyczaj używane w przypadku próbek linii zarodkowej, w przypadku których wiadomo, że ploidalności są dwie. Rozpoznawalne obszary referencyjne są najpierw identyfikowane z wystarczającym pokryciem dopasowania. W obrębie tych obszarów odniesienia szybkie skanowanie posortowanych odczytów identyfikuje obszary aktywne, które są wyśrodkowane na kolumnie nagromadzenia z dowodami wariantu. Te obszary aktywne są wyściełane kontekstem wystarczającym do pokrycia pobliskich znaczących treści bez odniesień. Jeśli są dowody na indele, obszary aktywne otrzymują dodatkową wyściółkę.

Dopasowane odczyty są przycinane w każdym aktywnym obszarze i łączone na wykresie De Bruijn. Krawędzie przyciętych odczytów są ważone według liczby obserwacji, a sekwencja odniesienia stanowi podstawę. Po pewnym oczyszczeniu i uproszczeniu wykresu wszystkie ścieżki źródło-odbiornik są wyodrębniane jako kandydujące haplotypy. Każdy haplotyp jest dopasowywany metodą Smitha-

Watermana do genomu referencyjnego w celu identyfikacji reprezentowanych przez niego wariantów. Ten zestaw zdarzeń może zostać uzupełniony przez wykrywanie oparte na położeniu. Dla każdej pary odczyt-haplotyp prawdopodobieństwo obserwacji odczytu  $P(r|H)$ , przy założeniu, że dany haplotyp jest rzeczywistą próbą początkową, jest szacowane przy użyciu pary ukrytego modelu Markowa (HMM).

Dzięki skanowaniu według pozycji referencyjnej na obszarze aktywnym, kandydujące genotypy są tworzone z diploidalnych kombinacji zdarzeń wariantowych (SNP lub indeli). Dla każdego zdarzenia (w tym odniesienia) warunkowe prawdopodobieństwo obserwacji  $P(r|e)$  każdego zachodzącego na siebie odczytu jest szacowane jako maksymalna wartość  $P(r|H)$  dla haplotypów uzasadniających zdarzenie. Są one łączone w warunkowe prawdopodobieństwo  $P(r|e1e2)$  dla genotypu (pary zdarzeń) i mnożone w celu uzyskania warunkowego prawdopodobieństwa  $P(R|e1e2)$  obserwacji całego nagromadzenia odczytów. Przy użyciu wzoru Bayesa oblicza się prawdopodobieństwo a posteriori  $P(e1e2|R)$  każdego genotypu diploidalnego, a zwycięzca zostaje wykryty.

DRAGEN for IDPE Dx stosuje automatyczne filtrowanie. Patrz [Adnotacje do pliku VCF procedury linii zarodkowej na stronie 13](#), aby uzyskać więcej informacji.

## Wykrywanie wariantu somatycznego DRAGEN

DRAGEN Germline Small Variant Caller przyjmuje zmapowane i wyrównane odczyty DNA jako dane wejściowe i rozpoznaje SNV i indele poprzez lokalny łączenie *de novo* haplotypów w obszarze aktywnym. Aby włączyć DRAGEN Somatic Small Variant Caller, wybierz aplikację wariantu somatycznego.

Rozpoznawanie wariantów somatycznych jest zwykle stosowane w przypadku próbek nowotworów. Dzięki tej procedurze DRAGEN nie przyjmuje żadnych założeń dotyczących ploidalności, co umożliwia wykrywanie alleli o niskiej częstotliwości. W przypadku loci z pokryciem do 100x w próbce guza, DRAGEN ma próg wykrywania przy częstotliwości wariantu alleli wynoszący 5%. Limit skaluje się wraz ze wzrostem głębokości na podstawie każdego locus i zmniejsza się o połowę za każdym razem, gdy zasięg podwaja się powyżej 100x. Rozpoznawalne obszary referencyjne są najpierw identyfikowane z wystarczającym pokryciem dopasowania. W obrębie tych obszarów odniesienia skanowanie posortowanych odczytów identyfikuje obszary aktywne, które są wyśrodkowane na kolumnach nagromadzenia z dowodami wariantu w odczytach nowotworu. Te obszary aktywne są wyścielane kontekstem wystarczającym do pokrycia pobliskich znaczących treści bez odniesień. Jeśli są dowody na indele, obszary aktywne otrzymują dodatkową wyściółkę.

Dopasowane odczyty są przycinane w każdym aktywnym obszarze i łączone na wykresie De Bruijn. Krawędzie przyciętych odczytów są ważone według liczby obserwacji, a sekwencja odniesienia stanowi podstawę. Po pewnym oczyszczeniu i uproszczeniu wykresu wszystkie ścieżki źródło-odbiornik są wyodrębniane jako kandydujące haplotypy. Każdy haplotyp jest dopasowywany metodą Smitha-Watermana do genomu referencyjnego w celu identyfikacji reprezentowanych przez niego wariantów. Dla każdej pary odczyt-haplotyp prawdopodobieństwo obserwacji odczytu  $P(r|H)$  szacuje się przy użyciu pary ukrytego modelu Markowa (HMM) przy założeniu, że haplotyp ten jest rzeczywistą próbą początkową.

W celu określenia wyniku granicy wykrywalności guza (TLOD), DRAGEN Somatic Small Variant Caller najpierw skanuje każde potencjalne zdarzenie somatyczne według pozycji referencyjnej, jak również zdarzenie referencyjne w aktywnym regionie. Warunkowe prawdopodobieństwo obserwacji  $P(r|e)$  każdego zachodzącego na siebie odczytu jest szacowane jako maksymalna wartość  $P(r|H)$  dla haplotypów uzasadniających zdarzenie. Są one łączone w warunkowe prawdopodobieństwo  $P(r|E)$  dla hipotezy zdarzenia,  $E$ , obejmujące mieszaninę alleli odniesienia i kandydata w zakresie możliwych częstotliwości alleli i mnożone w celu uzyskania warunkowego prawdopodobieństwa  $P(R|E)$  obserwacji całego nagromadzenia odczytów. Z tego poziomu obliczany jest wynik TLOD jako dowód na obecność allelu ALT w próbce guza w danym locus.

DRAGEN for IDPE Dx stosuje automatyczne filtrowanie. Patrz [Adnotacje plików VCF przepływu pracy somatycznej na stronie 16](#), aby uzyskać więcej informacji.



# Tworzenie zaplanowanego przebiegu

Aby skonfigurować przebieg w Illumina Run Manager na NextSeq 550Dx lub za pomocą przeglądarki na komputerze sieciowym, należy wykonać poniższe czynności. Jeśli wymagane jest zaimportowanie przykładowych danych, należy użyć przeglądarki na komputerze z dostępem do sieci. Instrukcje dotyczące uzyskiwania dostępu do Illumina Run Manager z komputera podłączonego do sieci znajdują się w Illumina Run Manager for NextSeq 550Dx – przewodnik po oprogramowaniu (dokument nr 200025239).

Istnieją dwa różne sposoby tworzenia nowego zaplanowanego przebiegu:

- **Importowanie przebiegu** – użyj arkusza próbek z istniejącego badania jako szablonu dla nowego badania. Informacje na temat importowania przebiegu zawiera Illumina Run Manager for NextSeq 550Dx – przewodnik po oprogramowaniu (dokument nr 200025239).
- **Utwórz przebieg** – wprowadź ręcznie parametry badania. Poniższe instrukcje opisują sposób tworzenia przebiegu.

**UWAGA** Pola wymagane w interfejsie użytkownika są oznaczone gwiazdką (\*).

## Application (Aplikacja)

1. Na karcie Planned (Zaplanowane) na ekranie Runs (Przebiegi) wybierz polecenie **Create Run** (Utwórz przebieg).
2. Wybierz aplikację DRAGEN for Illumina DNA Prep with Enrichment Dx, a następnie **Next** (Dalej).

## Ustawienia przebiegu

1. Na ekranie Run Settings (Ustawienia przebiegu) wprowadź unikalną nazwę przebiegu. Nazwa przebiegu to nazwa, która identyfikuje przebieg od sekwencjonowania po analizę.
2. **[Opcjonalnie]** Wprowadź opis przebiegu, aby dodatkowo ułatwić jego identyfikację.
3. Wybierz zestawy adaptera indeksującego używane podczas przygotowywania biblioteki.
4. Sprawdź długość odczytu i zmodyfikuj ją w razie potrzeby. Wartości domyślne Read 1 (Odczyt 1) i Read 2 (Odczyt 2) to 151 cykli. Index 1 (Indeks 1) i Index 2 (Indeks 2) mają stałą wartość 10 cykli i nie mogą być modyfikowane.
5. **[Opcjonalnie]** Wprowadź ID próbki Library.
6. Wybierz opcję **Next** (Dalej).

## Dane próbki

Dane próbki obejmują ID próbki, pozycję dołka (położenie dołka na płytce indeksowej) i nazwę biblioteki.

W przypadku korzystania z indeksu A&B pozycja dołka zawiera również identyfikator płytki.

Istnieją dwa sposoby wprowadzania danych próbek:

- **Import próbek** – użyj pliku szablonu dostępnego do pobrania na ekranie Sample Data (Dane próbki).
- **Ręcznie** – wprowadź dane próbki bezpośrednio do tabeli na ekranie Sample Data (Dane próbki).

### Importowanie próbek

Podczas planowania przebiegu sekwencjonowania przy użyciu przeglądarki na komputerze podłączonym do sieci, plik szablonu (\*.csv) jest dostępny do pobrania na ekranie Sample Data (Dane próbki). Plik szablonu nie jest dostępny do pobrania podczas uzyskiwania dostępu do Illumina Run Manager przez oprogramowanie systemu operacyjnego NextSeq 550Dx. Aby wprowadzić dane próbki za pomocą funkcji Import Samples (Import próbek), należy wykonać następujące czynności.

**UWAGA** Przed kontynuowaniem wykonaj kroki Run Settings (Ustawienia przebiegu).

1. Wybierz opcję **Download Template** (Pobierz szablon), aby pobrać pusty plik CSV.
2. Z pliku szablonu wprowadź przykładowe dane, a następnie zapisz plik. Library Name (Nazwa biblioteki) jest opcjonalna.

**UWAGA** W przypadku korzystania z indeksu A&B dane dla kolumny B muszą zawierać zarówno pozycję płytki, jak i dołka (pozycja dołka na płytce indeksowej). Przykład: A-A01, A-A02, A-A03.

3. Wybierz polecenie **Import Samples** (Importuj próbki) i przejdź do pliku szablonu zawierającego informacje o danych próbki z poprzedniego kroku.
4. Wybierz **Open** (Otwórz), **Proceed** (Kontynuuj), a następnie **Next** (Dalej).

**UWAGA** Zmiana Identyfikatora próbki przed wybraniem Next (Dalej) może spowodować błąd. Zakończ konfigurowanie przebiegu przed wprowadzeniem zmian, aby uniknąć błędów.

### Ręczne wprowadzanie próbek

Użyj tabeli na ekranie Sample Data (Dane próbki), aby ręcznie wprowadzić dane o próbce.

1. Wprowadź unikalny identyfikator próbki w polu Sample ID (Identyfikator próbki).
2. Użyj **Well Position** (Pozycja dołka) (Index A (Indeks A) lub Index B (Indeks B)) albo **Plate - Well Position** (Płytki – pozycja dołka) (Index A&B (Indeks A i B)), aby wybrać odpowiedni indeks dla próbek. Pola i7 Index, Index 1, i5 Index i Index 2 wypełniają się automatycznie.
3. **[Opcjonalne]** Wprowadź nazwę biblioteki.

4. Dodaj wiersze i powtarzaj kroki od **1** do **3**, według potrzeb, aż wszystkie próbki zostaną dodane do tej tabeli. Możesz dodać wiele wierszy jednocześnie, najpierw wprowadzając liczbę wierszy do dodania, a następnie wybierając ikonę +. Można również usunąć wiersze, zaznaczając pole obok numeru wiersza, a następnie klikając ikonę kosza.
5. Wybierz opcję **Next** (Dalej).

## Ustawienia analizy

1. Wybierz żądaną procedurę analizy:
  - Generowanie pliku FASTQ
  - Generowanie plików FASTQ i VCF dla procedury linii zarodkowej (wymagany plik manifestu)
  - Generowanie plików FASTQ i VCF dla procedury somatycznej (wymagany plik manifestu)
2. **[Opcjonalnie] Generate ORA compressed FASTQs** (Generowanie skompresowanych FASTQ ORA) jest domyślnie włączone. Kompresja FASTQ ORA bezstratnie kompresuje pliki FASTQ nawet 5 razy bardziej niż w przypadku fastq.gz. Usuń zaznaczenie **Generate ORA compressed FASTQs** (Generuj skompresowane ORA FASTQ), jeśli preferowane są nieskompresowane dane (fastq.gz).
3. W przypadku procedur linii zarodkowej i somatycznych wymagany jest plik wykazu. Użyj menu rozwijanego **Manifest File Selection** (Wybór pliku wykazu), aby wybrać plik wykazu. Wykaz jest plikiem BED (\*.bed) rozdzielanym tabulatorami, który określa nazwy i lokalizacje docelowych obszarów odniesienia. Więcej informacji można znaleźć w części [Plik wykazu na stronie 8](#) (Plik wykazu).
4. **[Opcjonalnie]**W przypadku procedur somatycznych użyj menu rozwijanego **Noise File Selection** (Wybór pliku szumu), aby wybrać systematyczny plik szumu. Można określić plik BED (\*.bed.gz) z poziomem szumu charakterystycznym dla danego miejsca w celu odfiltrowania systematycznego szumu. Więcej informacji podano w punkcie [Filtrowanie szumów \(opcjonalnie\) na stronie 9](#) (Filtrowanie szumu (opcja)).
5. Wybierz opcję **Next** (Dalej).

## Przebieg Przegląd

1. Na ekranie Review (Przegląd) wyświetlane są informacje dla Run Settings (Ustawienia przebiegu), Sample Data (Dane próbki) i Analysis Settings (Ustawienia analizy).
2. Wybierz **Save** (Zapisz).  
Przebieg zostanie zapisany na karcie Planned (Zaplanowane) na ekranie Runs (Przebiegi).

# Ustawienia

Aby wyświetlić lub zmienić ustawienia aplikacji DRAGEN for IDPE Dx, należy najpierw wybrać ikonę Applications (Aplikacje) na ekranie głównym. Następnie wybierz aplikację, którą chcesz wyświetlić lub zmienić. Do zmiany ustawień wymagane jest konto administratora.

## Konfiguracja

Na ekranie konfiguracji wyświetlane są następujące ustawienia aplikacji:

- **Library Prep Kits** – wyświetla domyślny zestaw do przygotowania biblioteki dla danej aplikacji. Tego ustawienia nie można zmienić.
- **Index Adapter Kits** – wyświetla domyślny zestaw adapterów indeksu dla danej aplikacji. Tego ustawienia nie można zmienić.
- **Read lengths** – długości odczytu są domyślnie ustawione na 151 dla danej aplikacji, ale można je zmienić podczas tworzenia przebiegu.
- **Manifest and Noise Files** – przesyłanie i zmiana ustawień dla plików wykazu i szumu.
  - Wybierz opcję **Upload File** (Prześlij plik), aby przesłać pliki do wykorzystania w analizie.
  - Wybierz przycisk opcji **Default** (Domyślne), aby ustawić plik jako domyślny plik wykazu lub plik szumu wybrany podczas tworzenia przebiegu po wybraniu danej aplikacji.
  - Zaznacz pole wyboru **Enabled** (Włączone), aby ustawić plik do wyświetlenia w menu rozwijanym podczas tworzenia przebiegu.

## Uprawnienia

Użyj pól wyboru na ekranie Permissions (Uprawnienia), aby zarządzać dostępem użytkowników do aplikacji.

## Plik wykazu

W przypadku korzystania z DRAGEN for IDPE Dx, wymagane jest wprowadzenie pliku wykazu dla następujących procedur:

- Generowanie FASTQ i VCF dla procedury linii zarodkowej
- Generowanie FASTQ i VCF dla procedury somatycznej

Plik wykazu jest tekstowym plikiem BED (\*.bed) rozdzielanym tabulatorami, który określa nazwy i lokalizacje docelowych obszarów odniesienia. Główną częścią pliku wykazu jest sekcja Regions (Regiony), która powinna zawierać następujące kolumny danych:

Kolumna	Opis
Name (Nazwa)	Unikatowa nazwa określona przez użytkownika dla celu
Chromosom	Lokalizacja chromosomu (np. chr10, chr5 itp.)
Start	indeks oparty na 1 dla pozycji początkowej celu
Dtop	indeks oparty na 1 dla pozycji końcowej celu
Długość sondy poprzedzającej	Długość sondy poprzedzającej. Dla aplikacji DRAGEN for IDPE Dx należy ustawić na 0.

Kolumna	Opis
Długość sondy następczej	Długość sondy następczej. Dla aplikacji DRAGEN for IDPE Dx należy ustawić na 0.

**UWAGA** Do analizy wymagany jest prawidłowy format pliku wykazu. DRAGEN zatrzyma analizę, jeśli plik wykazu będzie nieprawidłowy.

## Filtrowanie szumów (opcjonalnie)

Filtr szumu systematycznego jest dostępny dla rozpoznawania wariantów somatycznych i może być używany do redukcji fałszywie dodatnich rozpoznań poprzez uwzględnienie szumu charakterystycznego dla danego miejsca. Systematyczny plik szumu jest generowany przez zebranie około 50 normalnych próbek (najlepiej specyficznych dla panelu, przygotowania biblioteki i sekwensera), a następnie suma częstości alleli poniżej 30% w każdym miejscu z wystarczającym pokryciem jest dzielona przez całkowitą liczbę próbek (zakłada się, że częstości alleli powyżej 30% są wariantami linii zarodkowej, a nie szumem). Po wygenerowaniu wartości szumu warianty somatyczne wykryte w tym miejscu zostaną przefiltrowane.

Filtr ten może być używany w trybie Tumor-Normal (Guz-normalny), ale jest szczególnie przydatny w przebiegach Tumor-Only (Tylko guz), przy których nie jest dostępny dopasowany tryb normalny. Systematyczny plik szumów musi zawierać plik z rozszerzeniem BED (\* .bed.gz ) i musi zawierać cztery kolumny: Chromosome (Chromosom), Start (Początek), End (Koniec) i poziomy hałas specyficzne dla miejsca dla każdego wiersza. Systematyczne filtrowanie szumów jest opcjonalne.

## Dane wyjściowe analizy

Aktualnie trwające przebiegi wyświetlane są na karcie Active (Aktywne). Ukończone przebiegi są wyświetlane na karcie Completed (Ukończone). DRAGEN for IDPE Dx tworzy folder analizy o unikalnej nazwie dla każdej analizy, który jest oddzielony od folderu zawierającego dane sekwencjonowania. Folder analizy zawiera następujące informacje:

- Użyto pliku wykazu
- Wersja oprogramowania
- Identyfikatory próbek
- Łączna liczba dopasowanych odczytów
- Procent dopasowanych odczytów na próbkę
- Liczba rozpoznanych SNV na próbkę
- Liczba indeli rozpoznanych na próbkę
- Statystyki pokrycia

## Pliki wyjściowe analizy

Lokalizacja folderu analizy jest określona przez ustawienie External Storage for Analysis Results (Zewnętrzna pamięć dla wyników analizy). Więcej informacji na temat ustawienia External Storage for Analysis Results (Zewnętrzna pamięć dla wyników analizy) można znaleźć w Illumina Run Manager for NextSeq 550Dx – przewodnik po oprogramowaniu (dokument nr 200025239).

Na ekranie Run Details (Szczegóły przebiegu) pole External Location (Lokalizacja zewnętrzna) zawiera ścieżkę dla danych sekwencjonowania. Unikalna nazwa folderu analizy jest podana w polu Analysis Output Folder (Folder wyjściowy analizy) na ekranie Run Details (Szczegóły przebiegu). Dokładne wygenerowane pliki zależą od tego, którą procedurę analizy zastosowano. Następujące pliki wyjściowe analizy są generowane przez tę aplikację.

**UWAGA** Jeśli wystąpi błąd ograniczenia maksymalnej długości ścieżki pliku podczas uzyskiwania dostępu do plików wyjściowych analizy, spróbuj przenieść plik do lokalizacji o krótszej ścieżce lub użyj innej metody otwierania pliku.

Plik wyjściowy	Opis
Raport podsumowujący wariant (*.pdf)	Zawiera podsumowanie informacji o pliku, wersji oprogramowania, informacje o próbce, statystyki poziomu odczytu oraz SNV, insercje, delecje i podsumowania pokrycia. Tylko procedury dla linii zarodkowej i somatycznej tworzą raport wariantu.
FASTQ (*.fastq.gz lub *.fastq.ora)	Pliki pośrednie zawierające wyniki jakościowe rozpoznawania nukleotydów. Pliki FASTQ to podstawowe dane wejściowe etapu dopasowywania. Po wybraniu kompresji ORA używane jest rozszerzenie pliku *.fastq.ora.
Dopasowanie plików BAM (*.bam)	Zawiera dopasowane odczyty dla danej próbki.
Pliki VCF genomu (*.gvcf.gz)	Zawierają genotyp dla każdej pozycji, rozpoznany jako wariant lub materiał referencyjny.
Pliki VCF (*.vcf.gz)	Zawierają warianty rozpoznane w każdej pozycji.
Uruchamianie raportu wskaźników (*.csv)	Zawiera wskaźniki jakości dotyczące przebiegu, w tym nieindeksowany wynik uzysku całkowitego i Q30.

## Pliki FASTQ

FASTQ (\*.fastq.gz, \*.fastq.ora) to format pliku tekstowego, który zawiera rozpoznane nukleotydy i wartości dotyczące jakości dla każdego odczytu. Każdy plik zawiera następujące informacje:

- Identyfikator próbki
- Sekwencja

- Wyniki jakościowe w skali Phred w formacie kodowania ASCII + 33

Identyfikator próbki jest sformatowany w następujący sposób:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
Example:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<???'#=#
```

## Pliki BAM

Plik BAM (\*.bam) jest skompresowaną, binarną wersją pliku SAM (sequence alignment map – mapa dopasowania sekwencji), który służy do przechowywania odwzorowań dopasowanych sekwencji o wielkości do 128 Mb. W odniesieniu do plików BAM stosowany jest format nazwy `SampleName_S#.bam` (NazwaPróbki\_S#.bam), gdzie „#” jest numerem próbki wynikającym z kolejności, w jakiej próbki uszeregowano do przebiegu. W trybie wielowęzłowym numer S# jest ustawiony na S1, niezależnie od kolejności próbek.

Pliki BAM zawierają sekcje nagłówka i dopasowań:

- **Header** (Nagłówek) – zawiera informacje dotyczące całego pliku, takie jak nazwa próbki, długość próbki i metoda dopasowania. Elementy w sekcji dopasowań są powiązane z określonymi informacjami w sekcji nagłówka.
- **Alignments** (Dopasowania) – zawiera nazwę odczytu, sekwencję odczytu, jakość odczytu, informacje o dopasowaniu oraz znaczniki niestandardowe. Nazwa odczytu zawiera oznaczenia chromosomu, współrzędne początkowe, jakość dopasowania oraz ciąg deskryptora dopasowań.

Sekcja dopasowań zawiera następujące informacje dla każdego odczytu lub sparowanego odczytu:

- AS: jakość dopasowania odczytu w trybie sparowanych końców.
- RG: grupa odczytów wskazująca liczbę odczytów dla danej próbki.
- BC: znacznik kodu kreskowego, który wskazuje identyfikator demultipleksowanej próbki powiązany z danym odczytem.
- SM: jakość dopasowania odczytu w trybie pojedynczego końca.
- XC: ciąg deskryptora dopasowań.
- XN: znacznik nazwy ampliconu służący do zapisu identyfikatora ampliconu powiązanego z odczytem.

Pliki indeksów BAM (\*.bam.bai) zawierają indeks odpowiedniego pliku BAM.

## Pliki VCF

Pliki formatu rozpoznania wariantu (\*.vcf) zawierają informacje o wariantach znalezionych w określonych pozycjach w genomie referencyjnym.

Nagłówek pliku VCF zawiera wersję formatu pliku VCF i wersję algorytmu do rozpoznawania wariantów. Wymienione są w nim również adnotacje używane w pozostałej części pliku. Nagłówek VCF zawiera także plik genomu referencyjnego i plik BAM. Ostatni wiersz nagłówka zawiera nagłówki kolumny dla wierszy danych. Każdy wiersz danych pliku VCF zawiera informację o pojedynczym wariacie.

Tabela 1 Nagłówki pliku VCF

Nagłówek	Opis
CHROM (Chromosom)	Chromosom genomu referencyjnego. Chromosomy pojawiają się w takiej samej kolejności jak w referencyjnym pliku FASTA.
POS (Pozycja)	Pozycja pojedynczego nukleotydu w wariacie w chromosomie referencyjnym. W przypadku wariantów pojedynczych nukleotydów (SNV) ta pozycja jest nukleotydem referencyjnym dla danego wariantu. W przypadku indeli pozycja ta jest nukleotydem referencyjnym bezpośrednio poprzedzającym dany wariant.
ID (Identyfikator)	Numer rs (odniesienia SNP) dla SNP uzyskanego z pliku <code>dbSNP.txt</code> , jeśli ma to zastosowanie. Jeśli w danej lokalizacji jest wiele numerów rs, lista jest rozdzielona średnikami. Jeśli w danej pozycji nie ma żadnego wpisu dbSNP, stosuje się znacznik brakującej wartości ('.').
REF (Referencja)	Genotyp referencyjny. Przykładowo delekcja jednego nukleotydu T jest przedstawiana jako referencyjny allel TT i alternatywny T. Wariant pojedynczego nukleotydu A do T jest przedstawiany jako referencyjny allel A i alternatywny T.
ALT (Alternatywa)	Allele różniące się od odczytu referencyjnego. Na przykład: insercja jednego nukleotydu T jest przedstawiana jako referencyjny allel A i alternatywny AT. Wariant pojedynczego nukleotydu A do T jest przedstawiany jako referencyjny allel A i alternatywny T.
QUAL (Jakość)	Wynik jakościowy w skali Phred przypisany przez algorytm do rozpoznawania wariantów. Wyższe wyniki wskazują wyższy poziom ufności w odniesieniu do wariantu i niższe prawdopodobieństwo błędów. W przypadku wyniku jakościowego Q szacowane prawdopodobieństwo błędu wynosi $10^{-(Q/10)}$ . Na przykład zestaw rozpoznań z wynikiem Q30 ma odsetek błędów równy 0,1%. Wiele algorytmów rozpoznawania wariantów przypisuje wyniki jakościowe na podstawie swoich modeli statystycznych, które są wysokie w stosunku do obserwowanego odsetka błędów.



Tabela 2 Adnotacje do pliku VCF procedury linii zarodkowej

Nagłówek	Opis
FILTER (Filtr)	<p>Jeśli wariant przeszedł przez wszystkie filtry, w kolumnie filtru wpisywana jest informacja PASS (Powodzenie). Możliwe wpisy FILTERA obejmują:</p> <ul style="list-style-type: none"> <li>• <b>DRAGENSnpHardQUAL</b> – stosowany, jeśli wynik QUAL wariantu SNP nie spełnia progu</li> <li>• <b>DRAGENIndelHardQUAL</b> – stosowany, jeśli wynik QUAL wariantu indel nie spełnia progu</li> <li>• <b>LowDepth</b> – miejsce filtrowane, ponieważ głębokość pokrycia nie spełnia progu</li> <li>• <b>LowGQ</b> – miejsce filtrowane, ponieważ jakość genotypu nie spełnia progu</li> <li>• <b>PloidyConflict</b> – rozpoznanie genotypu z algorytmu rozpoznawania wariantów nie jest zgodne z ploidią chromosomu.</li> <li>• <b>base_quality</b> – miejsce filtrowane, ponieważ mediana jakości nukleotydów odczytów alternatywnych w tym locus nie spełnia progu</li> <li>• <b>filtered_reads</b> – miejsce filtrowane, ponieważ odfiltrowano zbyt dużą część odczytów</li> <li>• <b>fragment_length</b> – miejsce filtrowane, ponieważ różnica bezwzględna między medianą długości fragmentu odczytów alternatywnych a medianą długości fragmentów odczytów referencyjnych w tym locus przekracza próg</li> <li>• <b>low_depth</b> – miejsce filtrowane, ponieważ głębokość odczytu jest zbyt mała</li> <li>• <b>low_frac_info_reads</b> – miejsce filtrowane, ponieważ frakcja odczytów informatywnych jest poniżej progu</li> <li>• <b>low_normal_depth</b> – miejsce filtrowane, ponieważ normalna głębokość odczytu próbek jest zbyt mała</li> <li>• <b>long_indel</b> – miejsce filtrowane, ponieważ długość indela jest zbyt duża</li> <li>• <b>mapping_quality</b> – miejsce filtrowane, ponieważ mediana jakości odwzorowania odczytów alternatywnych w tym locus nie spełnia progu</li> <li>• <b>multiallelic</b> – miejsce filtrowane, ponieważ więcej niż dwa allele alternatywne przekraczają LOD nowotworu.</li> <li>• <b>non_homref_normal</b> – miejsce filtrowane, ponieważ genotyp próbki normalnej nie jest odniesieniem homozygotycznym</li> <li>• <b>no_reliable_supporting_read</b> – miejsce filtrowane, ponieważ nie ma wiarygodnego potwierdzającego odczytu somatycznego</li> <li>• <b>panel_of_normals</b> – obserwowany w co najmniej jednej próbce w panelu normalnych vcf</li> <li>• <b>read_position</b> – miejsce filtrowane, ponieważ mediana odległości między początkiem/końcem odczytu i tym locus jest poniżej progu</li> <li>• <b>RMxNRepeatRegion</b> – miejsce filtrowane, ponieważ całość lub część allelu tego wariantu jest powtórzeniem tego odniesienia</li> <li>• <b>strand_artifact</b> – miejsce filtrowane z powodu poważnego obciążenia systematycznego nici</li> <li>• <b>str_contraction</b> – miejsce filtrowane z powodu podejrzanego błędu PCR, gdzie alternatywny allel jest o jedną jednostkę powtórzenia mniejszy niż odniesienie</li> <li>• <b>too_few_supporting_reads</b> – miejsce filtrowane, ponieważ w próbce guza jest zbyt mało pomocniczych odczytów</li> <li>• <b>weak_evidence</b> – wynik wariantu somatycznego nie spełnia progu</li> </ul>

Nagłówek	Opis
INFO (Informacje)	<p data-bbox="347 254 735 285">Możliwe wpisy INFO obejmują:</p> <ul data-bbox="347 289 1481 1127" style="list-style-type: none"> <li data-bbox="347 289 1481 359">• <b>AC</b> – liczba alleli w genotypach dla każdego allelu ALT, w takiej samej kolejności, jak na liście</li> <li data-bbox="347 363 1386 394">• <b>AF</b> – częstość alleli dla każdego allelu ALT, w takiej samej kolejności, jak na liście</li> <li data-bbox="347 399 1038 430">• <b>AN</b> – łączna liczba alleli w rozpoznanych genotypach</li> <li data-bbox="347 434 639 466">• <b>DB</b> – udział w dBSNP</li> <li data-bbox="347 470 1453 548">• <b>FS</b> – wartość P skalowana według Phred przy użyciu dokładnego testu Fishera w celu wykrycia obciążenia systematycznego nici</li> <li data-bbox="347 552 1010 583">• <b>QD</b> – pewność/jakość wariantu według głębokości</li> <li data-bbox="347 588 1402 619">• <b>R2_5P_bias</b> – wynik oparty na obciążeniu mate i odległości od 5 głównych końców</li> <li data-bbox="347 623 1453 701">• <b>SOR</b> – symetryczny iloraz szans w tabeli kontyngencji 2x2 w celu wykrycia obciążenia systematycznego nici</li> <li data-bbox="347 705 1386 783">• <b>DP</b> – przybliżona głębokość odczytu (informatywna i nieinformatywna); niektóre odczyty mogły zostać odfiltrowane na podstawie mapq itp.</li> <li data-bbox="347 787 847 819">• <b>END</b> – pozycja zatrzymania interwału</li> <li data-bbox="347 823 1402 900">• <b>FractionInformativeReads</b> – udział odczytów informatywnych w całkowitej liczbie odczytów</li> <li data-bbox="347 905 751 936">• <b>MQ</b> – jakość mapowania RMS</li> <li data-bbox="347 940 1481 1018">• <b>MQRankSum</b> – Z-score w teście sumy rang Wilcoxona dla jakości mapowania odczytów alternatywnych w porównaniu z referencyjnymi</li> <li data-bbox="347 1022 1412 1100">• <b>ReadPosRankSum</b> – Z-score w teście sumy rang Wilcoxona dla obciążenia pozycji alternatywnych w porównaniu z referencyjnymi</li> <li data-bbox="347 1104 1227 1136">• <b>SOMATIC</b> – co najmniej jeden wariant w tej pozycji jest somatyczny</li> </ul>

Nagłówek	Opis
FORMAT (Format)	<p>W kolumnie formatu wymienione są pola rozdzielone dwukropkami. Przykładowo: GT:GQ</p> <p>Do dostępnych pól należą:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> – głębokości alleli (zliczanie tylko informatywnych odczytów z całkowitej liczby odczytów) dla alleli referencyjnych i alternatywnych w podanej kolejności</li> <li>• <b>AF</b> – frakcje alleli dla alleli alternatywnych w podanej kolejności</li> <li>• <b>DP</b> – przybliżona głębokość odczytu (odczyty z wartością MQ = 255 lub ze złym dopasowaniem par są odfiltrowywane)</li> <li>• <b>F1R2</b> – liczba odczytów w orientacji pary F1R2 potwierdzających każdy allel</li> <li>• <b>F2R1</b> – liczba odczytów w orientacji pary F2R1 potwierdzających każdy allel</li> <li>• <b>GT</b> – genotyp. 0 odpowiada nukleotydowi referencyjnemu, 1 odpowiada pierwszemu wpisowi w kolumnie ALT itd. Ukośnik (/) wskazuje brak informacji o fazowaniu.</li> <li>• <b>MB</b> – statystyki komponentów na próbkę w celu wykrycia obciążenia parowania</li> <li>• <b>PS</b> – informacje o identyfikatorze fazowania fizycznego, gdzie każdy unikatowy identyfikator w danej próbce (ale nie w wielu próbkach) łączy rekordy w grupie fazowania</li> <li>• <b>SB</b> – statystyki komponentów na próbkę, które składają się na dokładny test Fishera w celu wykrycia obciążenia systematycznego nici.</li> <li>• <b>SQ</b> – jakość somatyczna</li> </ul>
SAMPLE (Próbka)	<p>W kolumnie próbek są podane wartości wskazane w kolumnie FORMAT.</p>

Tabela 3 Adnotacje plików VCF przepływu pracy somatycznej

Nagłówek	Opis
FILTER (Filtr)	<p>Jeśli wariant przeszedł przez wszystkie filtry, w kolumnie filtru wpisywana jest informacja PASS (Powodzenie). Możliwe wpisy FILTERA obejmują:</p> <ul style="list-style-type: none"> <li>• <b>base_quality</b> – miejsce filtrowane, ponieważ mediana jakości nukleotydów odczytów alternatywnych w tym locus nie spełnia progu</li> <li>• <b>filtered_reads</b> – miejsce filtrowane, ponieważ odfiltrowano zbyt dużą frakcję odczytów</li> <li>• <b>fragment_length</b> – miejsce filtrowane, ponieważ różnica bezwzględna między medianą długości fragmentu odczytów alternatywnych a medianą długości fragmentów odczytów referencyjnych w tym locus przekracza próg</li> <li>• <b>low_depth</b> – miejsce filtrowane, ponieważ głębokość odczytu jest zbyt mała</li> <li>• <b>low_frac_info_reads</b> – miejsce filtrowane, ponieważ frakcja odczytów informatywnych jest poniżej progu</li> <li>• <b>low_normal_depth</b> – miejsce filtrowane, ponieważ normalna głębokość odczytu próbek jest zbyt mała</li> <li>• <b>long_indel</b> – miejsce filtrowane, ponieważ długość indela jest zbyt duża</li> <li>• <b>mapping_quality</b> – miejsce filtrowane, ponieważ mediana jakości odwzorowania odczytów alternatywnych w tym locus nie spełnia progu</li> <li>• <b>multiallelic</b> – miejsce filtrowane, ponieważ więcej niż dwa allele alternatywne przekraczają LOD nowotworu.</li> <li>• <b>non_homref_normal</b> – miejsce filtrowane, ponieważ genotyp próbki normalnej nie jest odniesieniem homozygotycznym</li> <li>• <b>no_reliable_supporting_read</b> – miejsce filtrowane, ponieważ nie ma wiarygodnego potwierdzającego odczytu somatycznego</li> <li>• <b>panel_of_normals</b> – obserwowany w co najmniej jednej próbce w panelu normalnych vcf</li> <li>• <b>read_position</b> – miejsce filtrowane, ponieważ mediana odległości między początkiem/końcem odczytu i tym locus jest poniżej progu</li> <li>• <b>RMxNRepeatRegion</b> – miejsce filtrowane, ponieważ całość lub część allelu tego wariantu jest powtórzeniem tego odniesienia</li> <li>• <b>strand_artifact</b> – miejsce filtrowane z powodu poważnego obciążenia systematycznego nici</li> <li>• <b>str_contraction</b> – miejsce filtrowane z powodu podejrzanego błędu PCR, gdzie alternatywny allel jest o jedną jednostkę powtórzenia mniejszy niż odniesienie</li> <li>• <b>too_few_supporting_reads</b> – miejsce filtrowane, ponieważ w próbce guza jest zbyt mało pomocniczych odczytów</li> <li>• <b>weak_evidence</b> – wynik wariantu somatycznego nie spełnia progu</li> <li>• <b>systematic_noise</b> – miejsce filtrowane na podstawie dowodów systematycznego szumu w wariantach normalnych</li> </ul>

Nagłówek	Opis
INFO (Informacje)	<p>Możliwe wpisy INFO obejmują:</p> <ul style="list-style-type: none"><li>• <b>DP</b> – przybliżona głębokość odczytu (informatywna i nieinformatywna); niektóre odczyty mogły zostać odfiltrowane na podstawie mapy itp.</li><li>• <b>END</b> – pozycja zatrzymania interwału</li><li>• <b>FractionInformativeReads</b> – udział odczytów informatywnych w całkowitej liczbie odczytów</li><li>• <b>MQ</b> – jakość mapowania RMS</li><li>• <b>MQRankSum</b> – Z-score w teście sumy rang Wilcoxa dla jakości mapowania odczytów alternatywnych w porównaniu z referencyjnymi</li><li>• <b>ReadPosRankSum</b> – Z-score w teście sumy rang Wilcoxa dla obciążenia pozycji alternatywnych w porównaniu z referencyjnymi</li><li>• <b>AQ</b> – wynik szumu systematycznego</li><li>• <b>hotspot</b> – znane miejsce somatyczne, używane do zwiększania wiarygodności rozpoznania</li><li>• <b>SOMATIC</b> – co najmniej jeden wariant w tej pozycji jest somatyczny</li></ul>

Nagłówek	Opis
FORMAT (Format)	<p>W kolumnie formatu wymienione są pola rozdzielone dwukropkami. Przykładowo: GT:GQ Do dostępnych pól należą:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> – głębokości alleli (zliczanie tylko informatywnych odczytów z całkowitej liczby odczytów) dla alleli referencyjnych i alternatywnych w podanej kolejności</li> <li>• <b>AF</b> – frakcje alleli dla alleli alternatywnych w podanej kolejności</li> <li>• <b>DP</b> – przybliżona głębokość odczytu (odczyty z wartością MQ = 255 lub ze złym dopasowaniem par są odfiltrowywane)</li> <li>• <b>F1R2</b> – liczba odczytów w orientacji pary F1R2 potwierdzających każdy allel</li> <li>• <b>F2R1</b> – liczba odczytów w orientacji pary F2R1 potwierdzających każdy allel</li> <li>• <b>GP</b> – prawdopodobieństwa a posteriori w skali Phred dla genotypów zgodnie ze specyfikacją VCF</li> <li>• <b>GQ</b> – jakość genotypu</li> <li>• <b>GT</b> – genotyp. 0 odpowiada nukleotydowi referencyjnemu, 1 odpowiada pierwszemu wpisowi w kolumnie ALT itd. Ukośnik (/) wskazuje brak informacji o fazowaniu.</li> <li>• <b>MB</b> – statystyki komponentów na próbkę w celu wykrycia obciążenia parowania</li> <li>• <b>PL</b> – znormalizowane prawdopodobieństwo genotypów w skali Phreda zdefiniowane w specyfikacji VCF</li> <li>• <b>PRI</b> – prawdopodobieństwo a priori genotypów w skali Phreda</li> <li>• <b>PS</b> – informacje o identyfikatorze fazowania fizycznego, gdzie każdy unikatowy identyfikator w danej próbce (ale nie w wielu próbkach) łączy rekordy w grupie fazowania</li> <li>• <b>SB</b> – statystyki komponentów na próbkę, które składają się na dokładny test Fishera w celu wykrycia obciążenia systematycznego nici.</li> <li>• <b>SQ</b> – jakość somatyczna</li> </ul>
SAMPLE (Próbka)	W kolumnie próbek są podane wartości wskazane w kolumnie FORMAT.

## Pliki VCF genomu

Pliki VCF genomu (\*.gvcf.gz) są zgodne ze zbiorem konwencji dotyczących przedstawienia wszystkich miejsc w genomie w dość kompaktowym formacie. Pliki gVCF zawierają wszystkie miejsca w obszarze docelowym w jednym pliku dla każdej próbki. W pliku gVCF w pozycjach, które nie przeszły wszystkich filtrów, przedstawione są nierozpoznane nukleotydy. Znacznik genotypu (GT) „./.” wskazuje nierozpoznany nukleotyd.

## Ponowne umieszczanie analizy w kolejce

Analizę można ponownie umieścić w kolejce po jej zatrzymaniu w przypadku jej niepowodzenia lub w celu ponownego przeanalizowania przebiegu przy różnych ustawieniach. Aby ponownie ustawić analizę w kolejce, wykonaj następujące czynności:

1. Na ekranie Run (Przebieg) wybierz kartę Completed (Ukończone), a następnie wybierz nazwę badania do ponownej analizy.  
Jeśli ponownie ustawienie analizy w kolejce zostało wcześniej przeprowadzone, wybierz nazwę przebiegu nadrzędnego.
2. Na ekranie Run Details (Szczegóły przebiegu), po Sequencing Information (informacje o sekwencjonowaniu), wybierz opcję **Requeue Analysis** (Ponownie umieść analizę w kolejce).
3. Wybierz opcję:
  - Requeue Analysis With No Changes (Ponownie umieść analizę w kolejce bez zmian)
  - Edytuj ustawienia przebiegu i ponownie umieść analizę w kolejce
  - Umieść analizę w kolejce z inną aplikacją
4. Potwierdź, że lokalizacja, w której obecnie znajdują się dane sekwencjonowania, jest podana w polu **Sequencing data file path** (Ścieżka do pliku danych sekwencjonowania).

**UWAGA** Ścieżka do danych sekwencjonowania powinna być zgodna ze ścieżką w ustawieniu External Storage for Analysis Results (Zewnętrzna pamięć dla wyników analizy). Informacje na temat zmiany zewnętrznej ścieżki przechowywania można znaleźć w Illumina Run Manager for NextSeq 550Dx – przewodnik po oprogramowaniu (dokument nr 200025239).

5. Wprowadź powód ponownej analizy.
6. Wybrać **Requeue Analysis** (Ponowne umieszczanie analizy w kolejce).
7. Edytuj żądane zmiany w Run Settings (Ustawienia przebiegu), Sample Data (Dane próbki) i Analysis Settings (Ustawienia analizy).
8. Wybierz **Save** (Zapisz). Analiza zostanie uruchomiona z jej bieżącymi parametrami.

## Pomoc techniczna

W celu uzyskania pomocy technicznej należy skontaktować się z działem pomocy technicznej firmy Illumina.

**Witryna:** [www.illumina.com](http://www.illumina.com)

**Adres e-mail:** [techsupport@illumina.com](mailto:techsupport@illumina.com)

**Karty charakterystyki (SDS)** – dostępne na stronie firmy Illumina pod adresem [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Dokumentacja produktu** jest dostępna do pobrania w witrynie [support.illumina.com](http://support.illumina.com).





Illumina  
5200 Illumina Way  
San Diego, California 92122, USA  
+1800 809 ILMN (4566)  
+1858 202 4566 (poza Ameryką Północną)  
techsupport@illumina.com  
www.illumina.com



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
The Netherlands

**Sponsor w Australii**

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

DO STOSOWANIA W DIAGNOSTYCE IN VITRO.

© 2023 Illumina, Inc. Wszelkie prawa zastrzeżone.

**illumina**<sup>®</sup>