

Lokalny menedżer przebiegu

Moduł analizy wariantów linii zarodkowej

Przewodnik dotyczący procedury działania w systemie
NextSeq 550Dx

DO STOSOWANIA W DIAGNOSTYCE IN VITRO

Przegląd	3
Wprowadzanie informacji o przebiegu	4
Metody analizy	6
Wyświetlanie danych dotyczących przebiegu i próbek	7
Raport z analizy	8
Pliki wyjściowe analizy	9
Historia wersji	16
Pomoc techniczna	17



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 mogą być wykorzystywane ani rozpowszechniane do innych celów 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.

© 2021 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 na stronie www.illumina.com/company/legal.html.

Przegląd

Moduł analizy wariantów linii zarodkowej lokalnego menedżera przebiegu jest przeznaczony do stosowania z zestawem Dx do niestandardowego sekwencjonowania amplikonów TruSeq firmy Illumina i aparatem NextSeq 550Dx. W przypadku stosowania z modułem analizy wariantów linii zarodkowej oznaczenie to jest przeznaczone do przygotowania bibliotek wykorzystywanych do sekwencjonowania DNA z próbek obwodowej krwi pełnej.

Moduł analizy ocenia krótkie regiony amplifikowanego DNA, czyli amplikony, pod kątem wariantów. Celowane sekwencjonowanie amplikonów zapewnia wysokie pokrycie określonych regionów.

Patrz: ulotka dołączona do opakowania zestawu Dx do niestandardowego sekwencjonowania amplikonów TruSeq (nr dokumentu: 100000029772).

Moduł analizy wariantów linii zarodkowej wymaga materiałów eksploatacyjnych na 300 cykli sekwencjonowania. Więcej informacji znajduje się w ulotce dołączonej do opakowania zestawu odczynników NextSeq 550Dx High Output Reagent Kit ver. 2 lub NextSeq 550Dx High Output Reagent Kit ver. 2.5.

Informacje na temat niniejszego przewodnika

W niniejszym przewodniku zawarto instrukcje dotyczące konfigurowania parametrów przebiegu sekwencjonowania i analizy przy użyciu modułu analizy wariantów linii zarodkowej. Informacje na temat panelu lokalnego menedżera przebiegu i ustawień systemu zawiera *Instrukcja obsługi aparatu NextSeq 550Dx* (nr dokumentu: 100000009513).

Lokalny menedżer przebiegu – przeglądanie

Interfejs lokalnego menedżera przebiegu jest wyświetlany w oprogramowaniu NextSeq 550Dx Operating Software (NOS) lub w przeglądarce internetowej. Obsługiwaną przeglądarką jest Chromium.



UWAGA

W przypadku używania nieobsługiwanej przeglądarki należy pobrać obsługiwaną przeglądarkę po wyświetleniu w oknie monitu „Confirm Unsupported Browser” (Potwierdź nieobsługiwaną przeglądarkę). Wybrać „**here**” (tutaj), aby pobrać obsługiwaną wersję przeglądarki Chromium.

Wyświetlanie na monitorze aparatu

- 1 Aby wyświetlić interfejs lokalnego menedżera przebiegu na monitorze aparatu, należy wybrać jedną z następujących opcji:
 - ▶ Na ekranie głównym oprogramowania NOS wybrać opcję **Local Run Manager** (Lokalny menedżer przebiegu).
Kliknąć X w prawym górnym rogu, aby po zakończeniu wrócić do ekranu oprogramowania NOS.
 - ▶ Wybrać ikonę minimalizacji NOS, otworzyć przeglądarkę Chromium na aparacie, a następnie wpisać **http://localhost** na pasku adresu.
Tylko administratorzy mogą zminimalizować oprogramowanie NOS.

Wyświetlanie na komputerze sieciowym

- 1 Otworzyć przeglądarkę internetową Chromium na komputerze z dostępem do tej samej sieci co aparat i połączyć się za pomocą adresu IP aparatu lub nazwy aparatu. Na przykład **http://myinstrument**.

Wprowadzanie informacji o przebiegu

Ustawianie parametrów

- 1 Zalogować się do lokalnego menedżera przebiegu.
- 2 Wybrać opcję **Create Run** (Utwórz przebieg), a następnie opcję **Germline Variant** (Wariant linii zarodkowej).
- 3 Wprowadzić nazwę przebiegu, która identyfikuje przebieg od sekwencjonowania po analizę. Można używać znaków alfanumerycznych, spacji, znaków podkreślenia lub łączników.
- 4 **[Opcjonalnie]** Wprowadzić opis przebiegu, aby ułatwić jego identyfikację. Można używać znaków alfanumerycznych, spacji, znaków podkreślenia lub łączników.
- 5 Wybrać liczbę próbek i zbiór indeksów z listy rozwijanej. Dokonując wyboru, należy uwzględnić poniższe informacje.
 - ▶ Lista rozwijana zawiera liczby próbek ze zbiorem indeksów. Na przykład „24-Set 1” oznacza 24 próbki do przetestowania z indeksami ze zbioru 1 indeksów.
 - ▶ Numery zbiorów indeksów odnoszą się do różnych zbiorów indeksów i5. Oba zbiory Set 1 (Zbiór 1) i Set 2 (Zbiór 2) zapewniają zróżnicowanie indeksów. Proponowane są dwa zbiory indeksów, aby zapobiec wyczerpaniu pojedynczego zbioru.
 - ▶ Należy wybrać liczbę próbek najbliższą liczbie próbek do testowania. Jeśli dokładnej liczby próbek nie ma na liście, należy wybrać najbliższą wartość, jednak mniejszą od liczby testowanych próbek. Na przykład: jeśli chcesz przetestować 18 próbek, wybierz liczbę 16.
 - ▶ Kombinacje dołków próbek i indeksów spełniające wymagania dotyczące zróżnicowania są zaznaczone na zielono. W przypadku wybrania innych kombinacji dołków i indeksów, jeśli wymagania dotyczące zróżnicowania indeksów nie są spełnione, podczas zapisywania przebiegu pojawi się powiadomienie.

Importowanie plików wykazów dla przebiegu

- 1 Upewnić się, że wykazy, które mają być zaimportowane, znajdują się w dostępnej lokalizacji sieciowej lub na dysku USB.
- 2 Wybrać opcję **Import Manifests** (Importuj wykazy).
- 3 Przejść do pliku wykazu i wybrać wykazy do dodania.



UWAGA

Aby udostępnić pliki wykazów dla wszystkich przebiegów przy użyciu modułu analizy wariantów linii zarodkowej, należy dodać wykazy za pomocą funkcji Module Settings (Ustawienia modułu). Do jej użycia wymagane są uprawnienia administratora. Więcej informacji na ten temat zawiera *Instrukcja obsługi aparatu NextSeq 550Dx* (nr dokumentu: 1000000009513).

Określanie próbek do przebiegu


Próbki do przebiegu określa się za pomocą jednej z poniższych opcji, zgodnie z podanymi wskazówkami.

- ▶ **Enter samples manually** (Ręczne wprowadzanie próbek) – należy użyć pustej tabeli na ekranie Create Run (Tworzenie przebiegu).


- ▶ **Import samples** (Importowanie próbek) – należy przejść do pliku zewnętrznego w formacie wartości rozdzielonych przecinkami (*.csv). Szablon jest dostępny do pobrania na ekranie Create Run (Tworzenie przebiegu).

Po wypełnieniu tabeli próbek można wyeksportować informacje o próbce do pliku zewnętrznego. Plik ten posłuży jako odniesienie podczas przygotowywania bibliotek lub importowania pliku na potrzeby kolejnego przebiegu.

Ręczne wprowadzanie próbek

- 1 Wprowadzić niepowtarzalną nazwę próbki w polu Sample ID (Identyfikator próbki). Można używać znaków alfanumerycznych, łączników lub znaków podkreślenia.
- 2 **[Opcjonalnie]** W przypadku dodatnich lub ujemnych próbek kontrolnych kliknąć prawym przyciskiem myszy i wybrać typ próbki kontrolnej.
- 3 **[Opcjonalnie]** Wprowadzić opis próbki w polu Sample Description (Opis próbki). Można używać znaków alfanumerycznych, łączników lub znaków podkreślenia. Opisy próbki są powiązane z nazwą próbki. Jeśli ta sama nazwa próbki zostanie ponownie użyta w późniejszym przebiegu, opisy próbki zostaną nadpisane.
- 4 Wybrać adapter Indeksu 1 z listy rozwijanej Index 1 (i7) (Indeks 1, i7). W przypadku użycia sugerowanych dołków próbek oprogramowanie automatycznie wypełnia adaptory indeksu i7 oraz i5, które spełniają wymogi dotyczące zróżnicowania indeksu. Jeśli na liście nie ma dokładnej liczby testowanych próbek, należy wybrać adaptory indeksu dla dodatkowych dołków. Jeśli istnieje potrzeba, aby wybrać indeksy dla dodatkowych dołków, lub nie będą używane zalecane kombinacje adapterów indeksu, przed wybraniem indeksów należy przeczytać część *Rozpoznawanie nukleotydów i zróżnicowanie indeksu na stronie 15*.
- 5 Wybrać adapter Indeksu 2 z listy rozwijanej Index 2 (i5) (Indeks 2, i5).
- 6 Wybrać plik wykazu z listy rozwijanej Manifest (Wykaz).
- 7 Wybrać odpowiednią opcję, aby wyświetlić, wydrukować lub zapisać układ płytki jako odniesienie do przygotowywania bibliotek.
 - ▶ Wybrać ikonę  **Drukuj**, aby wyświetlić układ płytki. Wybrać opcję **Print** (Drukuj), aby wydrukować układ płytki.
 - ▶ Wybrać opcję **Export** (Eksportuj), aby wyeksportować informacje o próbce do pliku zewnętrznego. Upewnić się, że wykaz i informacje o próbce są prawidłowe. Nieprawidłowe informacje mogą wpłynąć na wyniki.
- 8 Wybrać opcję **Save Run** (Zapisz przebieg).

Importowanie próbek

- 1 Wybrać opcję **Import Samples** (Importuj próbki) i przejść do lokalizacji pliku z informacjami o próbce. Można zaimportować dwa rodzaje plików.
 - ▶ Wybrać opcję **Template** (Szablon) na ekranie Create Run (Tworzenie przebiegu), aby przygotować nowy układ płytki. Plik szablonu zawiera właściwe nagłówki kolumn do zaimportowania. Dla próbek w przebiegu w każdej kolumnie wprowadzić informacje o próbce. Usunąć przykładowe informacje w nieużywanych komórkach, a następnie zapisać plik.
 - ▶ Użyć pliku z informacjami o próbce, który został wyeksportowany z modułu analizy wariantów linii zarodkowej przy użyciu funkcji eksportowania.
- 2 Wybrać ikonę  **Drukuj**, aby wyświetlić układ płytki.

- 3 Wybrać opcję **Print** (Drukuj), aby wydrukować układ płytki jako odniesienie do przygotowania bibliotek.
- 4 **[Opcjonalnie]** Wybrać opcję **Export** (Eksportuj), aby wyeksportować informacje o próbce do pliku zewnętrznego.
Upewnić się, że wykaz i informacje o próbce są prawidłowe. Nieprawidłowe informacje mogą wpłynąć na wyniki.
- 5 Wybrać opcję **Save Run** (Zapisz przebieg).

Edytowanie przebiegu

Instrukcje na temat edytowania informacji dotyczących przebiegu przed rozpoczęciem sekwencjonowania zawiera *Instrukcja obsługi aparatu NextSeq 550Dx (nr dokumentu: 100000009513)*.

Metody analizy

Moduł analizy wariantów linii zarodkowej wykonuje wymienione poniżej etapy analizy, a następnie zapisuje pliki wyjściowe analizy w folderze Alignment (Dopasowanie).

- ▶ Demultipleksowanie odczytów indeksów
- ▶ Generowanie plików FASTQ
- ▶ Dopasowanie do materiału referencyjnego
- ▶ Identyfikacja wariantów

Demultipleksowanie

Demultipleksowanie polega na porównywaniu każdej sekwencji odczytu indeksów z sekwencjami indeksów określonymi dla przebiegu. Na tym etapie nie są brane pod uwagę żadne wartości dotyczące jakości.

Odczyty indeksów identyfikuje się poprzez następujące czynności:

- ▶ Próbki są ponumerowane, zaczynając 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 jest dokładnie zgodna lub gdy występuje maksymalnie jedna niezgodność na odczyt indeksu.

Generowanie pliku FASTQ

Po zakończeniu demultipleksowania oprogramowanie generuje pliki analizy pośredniej w formacie FASTQ, który przedstawia sekwencje w postaci tekstowej. Pliki FASTQ zawierają odczyty dla każdej próbki i powiązane z nimi wyniki jakościowe. Klastry, które nie przeszły przez filtr, są wykluczane.

Każdy plik FASTQ zawiera odczyty tylko dla jednej próbki. Nazwa tej próbki jest uwzględniona w nazwie pliku FASTQ. Pliki FASTQ to podstawowe dane wejściowe do dopasowywania. Dla każdej próbki generowanych jest osiem plików FASTQ – cztery z Odczytu 1 i cztery z Odczytu 2.

Dopasowywanie

Podczas etapu dopasowywania klastry z poszczególnych próbek są dopasowywane za pomocą pasmowego algorytmu Smitha-Watermana do sekwencji ampliconu określonych w pliku wykazu.

Pasmowy algorytm Smitha-Watermana wykonuje półglobalne dopasowania sekwencji w celu określenia podobnych regionów między dwiema sekwencjami. Zamiast porównywać całą sekwencję, algorytm Smitha-Watermana porównuje segmenty o wszystkich możliwych długościach.

Każdy odczyt w trybie sparowanych końców jest oceniany pod względem dopasowania do sekwencji sondy odpowiednich dla tego odczytu.

- ▶ Odczyt 1 jest oceniany względem odwrotnej zgodności oligonukleotydów dalszych zależnych od umiejscowienia (ang. Downstream Locus-Specific Oligos, DLSSO).
- ▶ Odczyt 2 jest oceniany względem zgodności oligonukleotydów wcześniejszych zależnych od umiejscowienia (ang. Upstream Locus-Specific Oligos, ULSSO).
- ▶ Jeśli przy dopasowaniu początku odczytu do sekwencji sondy występują maksymalnie trzy różnice (niedopasowania lub przesunięcia z powodu wiodących polimorfizmów typu indel), pełna długość odczytu jest dopasowywana do amplikonu docelowego dla tej sekwencji.
- ▶ Ze względu na mechanizm działania oznaczenia polimorfizmy typu indel w oligonukleotydach DLSSO i ULSSO nie są obserwowane.

Dopasowania są odfiltrowywane z wyników na podstawie liczby niedopasowań w regionie docelowym lub całym amplikonie, w zależności od długości amplikonu. Odfiltrowane dopasowania są zapisywane w plikach dopasowania jako niedopasowane i nie są używane do rozpoznawania wariantów.

Rozpoznawanie wariantów

Opracowany przez firmę Illumina algorytm do rozpoznawania wariantów PISCES identyfikuje warianty występujące w próbce DNA.



Algorytm do rozpoznawania wariantów PISCES identyfikuje warianty SNV, warianty MNV i małe polimorfizmy typu indel poprzez wykonanie 3 etapów:

- ▶ Oddzielne rozważenie każdej pozycji w genomie referencyjnym.
- ▶ Zliczenie nukleotydów w określonych pozycjach pod kątem dopasowanych odczytów nakładających się na tę pozycję.
- ▶ Obliczenie wyniku dla wariantu, który określa jakość rozpoznania przy użyciu modelu Poissona. Warianty z wynikiem jakościowym poniżej Q20 są wykluczane.

Jeśli wariant przejdzie wszystkie filtry, zostaje oznaczony jako PASS (Powodzenie) w VCF.

Więcej informacji można znaleźć na stronie github.com/Illumina/PISCES/wiki.

Wyświetlanie danych dotyczących przebiegu i próbek

- 1 Kliknąć nazwę przebiegu w panelu lokalnego menedżera przebiegu.
- 2 Przejrzeć pomiary sekwencjonowania na karcie Run Overview (Przegląd przebiegu).
- 3 **[Opcjonalnie]** Kliknąć ikonę **Kopij do schowka** , aby skopiować ścieżkę do folderu wyjściowego przebiegu.
- 4 Kliknąć kartę Sequencing Information (Informacje dotyczące sekwencjonowania), aby przejrzeć parametry przebiegu i informacje o materiałach eksploatacyjnych.
- 5 Kliknąć kartę Samples and Results (Próbki i wyniki), aby wyświetlić lokalizację raportu z analizy.
 - ▶ Jeśli analizę powtarzano, rozwinąć listę Select Analysis (Wybierz analizę) i wybrać odpowiednią analizę.
- 6 Kliknąć ikonę **Kopij do schowka** , aby skopiować ścieżkę do folderu Analysis (Analiza).

Więcej informacji na temat kart Run Overview (Przegląd przebiegu) i Sequencing Information (Informacje dotyczące sekwencjonowania) oraz sposobu ponownego umieszczania analizy w kolejce zawiera *Instrukcja obsługi aparatu NextSeq 550Dx* (nr dokumentu: 100000009513).

Raport z analizy

Wyniki analizy są podsumowane na karcie Samples and Results (Próbki i wyniki) oraz w postaci raportu zbiorczego w folderze Alignment (Dopasowanie). Raport dla każdej próbki jest też dostępny w formacie pliku PDF.

Informacje na karcie Samples and Results (Próbki i wyniki)

1 Kliknij próbkę na liście, aby zobaczyć raport dotyczący próbki.

Tabela 1 Informacje o przebiegu i próbce

Nagłówek kolumny	Opis
Run Status (Stan przebiegu)	Określa, czy sekwencjonowanie zakończyło się powodzeniem czy niepowodzeniem.
Total Yield (GB) (Uzysk całkowity, GB)	Liczba nukleotydów rozpoznanych poprzez sekwencjonowanie. Przedstawia próg powodzenia oraz stan: powodzenie lub niepowodzenie.
% ≥ Q30	Procent odczytów podczas sekwencjonowania z wynikiem jakościowym równym 30 (Q30) lub wyższym. Przedstawia próg powodzenia oraz stan: powodzenie lub niepowodzenie.
Sample ID (Identyfikator próbki)	Identyfikator próbki podany w momencie tworzenia przebiegu.
Total PF Reads (Odczyty PF łącznie)	Łączna liczba odczytów, które przeszły przez filtr.
Read 1 % ≥ Q30 (% Odczytu 1 ≥ Q30)	Odsetek odczytów w Odczycie 1 z wynikiem jakościowym równym 30 (Q30) lub wyższym dla próbki.
Read 2 % ≥ Q30 (% Odczytu 2 ≥ Q30)	Odsetek odczytów w Odczycie 2 z wynikiem jakościowym równym 30 (Q30) lub wyższym dla próbki.
Autosome Call Rate (Wskaźnik rozpoznań autosomów)	Liczba pozycji genomowych w obrębie autosomów (chromosomów od 1 do 22), które spełniają zdefiniowany wstępnie próg wartości ufności, podzielona przez całkowitą liczbę przeanalizowanych autosomalnych pozycji genomowych. Wskaźnik rozpoznań jest opisywany w odniesieniu do próbki i zgłaszany jako odsetek obliczany według wzoru: 1 minus (liczba pozycji autosomalnych z niepełnymi rozpoznaniem) podzielona przez łączną liczbę sekwencjonowanych pozycji autosomalnych).

Tabela 2 Informacje w raporcie dotyczącym próbki

Nagłówek kolumny	Opis
Sample (Próbka)	Identyfikator próbki podany w momencie tworzenia przebiegu.
Report Date (Data raportu)	Data wygenerowania raportu.
Sample Information (Informacje o próbce)	Identyfikator próbki podany w momencie tworzenia przebiegu, łączna liczba odczytów w próbce, które przeszły przez filtr, odsetek odczytów dla próbki o wyniku jakościowym równym co najmniej 30 (Q30) i wskaźnik rozpoznań autosomalnych.
Amplicon Summary (Podsumowanie amplikonu)	Łączna liczba sekwencjonowanych regionów amplikonu oraz całkowita długość par nukleotydów sekwencjonowanych amplikonów w regionach docelowych dla próbki oraz pliku wykazu. Plik wykazu określa genom referencyjny i docelowe regiony referencyjne używane na etapie dopasowywania.
Read Level Statistics (Statystyki poziomu odczytu)	Liczba i procent odczytów dla próbki z pokryciem każdej pozycji w materiale referencyjnym dla Odczytu 1 i Odczytu 2.

Nagłówek kolumny	Opis
Variants Summary (Podsumowanie wariantów)	Liczba wariantów pojedynczych nukleotydów (SNV), insercji i delecji wykrytych dla próbki, które są zgodne z wartościami sugerowanymi w celu ustalenia, czy wyniki jakościowe nie przekraczają dopuszczalnego zakresu.
Coverage Summary (Podsumowanie pokrycia)	Łączna liczba dopasowanych nukleotydów podzielona przez rozmiar regionu docelowego oraz odsetek regionów ampliconu z wartościami pokrycia powyżej dolnego progu pokrycia wynoszącego $0,2 \times$ średnie pokrycie ampliconu dla próbki.
Coverage Plots (Wykresy pokrycia)	Wykres pokrycia według regionu ampliconu przedstawia pokrycie w obrębie regionów ampliconu dla próbki. Regiony z wartościami pokrycia niższymi od progu pokrycia są zaznaczone na czerwono. Średnia wszystkich wartości jest wskazana przez pomarańczową linię.
Software Versions (Wersje oprogramowania)	Wersje oprogramowania używane podczas sekwencjonowania próbki. Obejmują wersję oprogramowania sterującego aparatem NextSeq 550Dx (NOS), oprogramowania lokalnego menedżera przebiegu, oprogramowania RTA i modułu analizy wariantów linii zarodkowej.

Pliki wyjściowe analizy

Na potrzeby modułu analizy wariantów linii zarodkowej generowane są wymienione poniżej pliki wyjściowe analizy. Zawierają one wyniki analizy potrzebne do dopasowywania i rozpoznawania wariantów. Pliki wyjściowe analizy znajdują się w folderze Alignment (Dopasowanie).

Nazwa pliku	Opis
Demultipleksowanie (*.txt)	Pliki pośrednie zawierające podsumowanie wyników demultipleksowania.
FASTQ (*.fastq.gz)	Pliki pośrednie zawierające wyniki jakościowe rozpoznawania nukleotydów. Pliki FASTQ to podstawowe dane wejściowe etapu dopasowywania.
Pliki dopasowywania w formacie BAM (*.bam)	Zawierają dopasowane odczyty dla danej próbki.
Pliki rozpoznawania wariantów w formacie VCF genomu (*.genome.vcf.gz)	Zawierają genotyp dla każdej pozycji, rozpoznany jako wariant lub materiał referencyjny.
Pliki rozpoznawania wariantów w formacie VCF (*.vcf.gz)	Zawierają wszystkie warianty rozpoznane w obszarze docelowym.
AmpliconCoverage_M1.tsv	Zawiera informacje o pokryciu poszczególnych ampliconów na próbkę według każdego podanego wykazu. Litera M z liczbą oznacza numer wykazu.

Format pliku demultipleksowania

W procesie demultipleksowania następuje odczyt sekwencji indeksu dołączonej do każdego klastra, aby określić, z której próbki pochodzi klastery. Mapowanie między klastrami i numerem próbki jest zapisywane w pliku demultipleksowania (*.demux) dla każdej płytki komory przepływowej.

Format nazwy pliku demultipleksowania to **s_1_X.demux**, gdzie „X” to numer płytki.

Plik demultipleksowania zaczyna się od nagłówka:

- ▶ Wersja (4-bajtowa liczba całkowita), obecnie: 1
- ▶ Liczba klastrów (4-bajtowa liczba całkowita)

Pozostała część pliku składa się z numerów próbek dla każdego klastra z płytki.

Po zakończeniu etapu demultipleksowania oprogramowanie generuje plik demultipleksowania o nazwie **DemultiplexSummaryF1L1.txt**.

- ▶ **F1** w nazwie pliku odpowiada numerowi komory przepływowej.
- ▶ **L1** w nazwie pliku odpowiada numerowi pasma.
- ▶ Wyniki demultipleksowania przedstawione są w tabeli z 1 wierszem na płytkę i 1 kolumną na próbkę, z uwzględnieniem próbki 0.
- ▶ Najczęściej występujące sekwencje w odczytach indeksów.

Format pliku FASTQ

FASTQ to format pliku tekstowego, który zawiera rozpoznane nukleotydy i wartości dotyczące jakości dla każdego odczytu. Każdy rekord zawiera 4 wiersze:

- ▶ Identyfikator
- ▶ Sekwencja
- ▶ Znak plus (+)
- ▶ Wyniki jakościowe w skali Phred w formacie kodowania ASCII + 33

Identyfikator ma następujący format:

@Aparat:IDprzebiegu:IDKomoryPrzepływowej:Pasma:Płytką:X:Y NrOdczytu:FlagaFiltra:0:NumerPróbki

Przykład:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

Format pliku BAM

Plik BAM (*.bam) jest skompresowaną, binarną wersją pliku SAM, który służy do przechowywania odwzorowań dopasowanych sekwencji. Rozmiar pliku wynosi maksymalnie 128 Mb. Formaty SAM i BAM opisano szczegółowo w dokumencie samtools.github.io/hts-specs/SAMv1.pdf.

W odniesieniu do plików BAM stosowany jest format nazwy **NazwaPróbki_S#.bam**, gdzie „#” jest numerem próbki wynikającym z kolejności, w jakiej wymieniono próbki dla przebiegu.

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 odczytu wskazująca liczbę odczytów dla danej próbki.
- ▶ **BC:** znacznik kodu kreskowego, który wskazuje identyfikator demultipleksowanej próbki powiązany z 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.

Format pliku VCF

VCF (ang. Variant Call Format) to powszechnie używany format pliku opracowany przez społeczność naukową zajmującą się genomiką. Plik zawiera informacje o wariantach znalezionych w konkretnych pozycjach w genomie referencyjnym. Pliki VCF mają rozszerzenie .vcf.

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.

Nagłówki pliku VCF

Nagłówek	Opis
CHROM (Chromosom)	Chromosom genomu referencyjnego. Chromosomy pojawiają się w takiej samej kolejności jak w referencyjnym pliku FASTQ.
POS (Pozycja)	Pozycja pojedynczego nukleotydu w wariacie w chromosomie referencyjnym. W przypadku polimorfizmów SNP ta pozycja to nukleotyd referencyjny z wariantem; w przypadku polimorfizmów typu indel lub delecji ta pozycja to nukleotyd referencyjny bezpośrednio przed wariantem.
ID (Identyfikator)	Numer rs wariantu uzyskany z pliku dbSNP.txt, jeśli ma to zastosowanie. Jeśli w danej lokalizacji istnieje wiele numerów rs, lista jest rozdzielona średnikami. Jeśli w danej pozycji nie istnieje żaden wpis bazy danych dbSNP, zostanie zastosowany znacznik brakującej wartości ('.').
REF (Referencja)	Genotyp referencyjny. Przykładowo delecja 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.

Adnotacje w pliku VCF

Nagłówek	Opis
FILTER (Filtr)	<p>Jeśli wariant przeszedł przez wszystkie filtry, w kolumnie filtra wpisywana jest informacja PASS (Powodzenie).</p> <ul style="list-style-type: none"> • LowDP (Mała głębokość) — dotyczy miejsc z głębokością pokrycia poniżej 150x. W przypadku pozycji ampliconu pokrytych zarówno przez odczyt puli oligonukleotydów przednich i wstecznych jest to odpowiednik 300 nakładających się odczytów w trybie sparowanych końców. • q20 (Wynik jakościowy) – wynik jakościowy < 20. • MultiAllelicSite (Miejsce wieloalleliczne) – wariant niezgodny z modelem diploidalnym. • R5x9 – liczba sąsiednich powtórzeń (o długości od 1 do 5 bp) do rozpoznania wariantów ≥ 9. • SB – obciążenie systematyczne nici przekracza podany próg.
INFO (Informacje)	<p>Do możliwych wpisów w kolumnie INFO (Informacje) należą:</p> <ul style="list-style-type: none"> • AC – liczba alleli w genotypach dla każdego allelu ALT, w takiej samej kolejności jak na liście. • AF – częstość alleli dla każdego allelu ALT, w takiej samej kolejności jak na liście. • AN – łączna liczba alleli w rozpoznanych genotypach. • CD – flaga wskazująca, że SNP występuje w regionie kodującym co najmniej 1 wpisu bazy RefGene. • DP – głębokość (liczba rozpoznanych nukleotydów dopasowanych do pozycji i użytych w rozpoznawaniu wariantów). • Exon – rozdzielona przecinkami lista regionów eksonów odczytana z bazy RefGene. • FC – funkcjonalna konsekwencja. • GI – rozdzielona przecinkami lista identyfikatorów genów odczytana z bazy RefGene. • QD – pewność/jakość wariantu według głębokości. • TI – rozdzielona przecinkami lista identyfikatorów transkryptów odczytana z bazy RefGene.
FORMAT (Format)	<p>W kolumnie formatu wymienione są pola rozdzielone dwukropkami. Przykładowo: GT:GQ. Lista podanych pól zależy od użytego algorytmu do rozpoznawania wariantów. Do dostępnych pól należą:</p> <ul style="list-style-type: none"> • AD – wpis w postaci „X,Y”, gdzie X jest liczbą rozpoznanych referencyjnych, a Y jest liczbą rozpoznanych alternatywnych. • DP – przybliżona głębokość odczytu; odczyty z wartością MQ = 255 lub ze złym dopasowaniem par są odfiltrowywane. • GQ – jakość genotypu. • GT – genotyp. 0 odpowiada nukleotydowi referencyjnemu, 1 odpowiada pierwszemu wpisowi w kolumnie ALT itd. Ukośnik (/) wskazuje brak informacji o fazowaniu. • NC – frakcja nukleotydów, które nie zostały rozpoznane lub o jakości rozpoznawania nukleotydów poniżej minimalnej wartości progowej. • NL – poziom szumu; szacowany szum przy rozpoznawaniu nukleotydów w tej pozycji. • SB – obciążenie systematyczne nici w tej pozycji. Większe wartości ujemne wskazują mniejsze obciążenie systematyczne; wartości zbliżone do 0 wskazują większe obciążenie. • VF – częstość wariantów; procent odczytów wspierających alternatywny allel.
SAMPLE (Próbka)	W kolumnie próbek są podane wartości wskazane w kolumnie FORMAT.

Pliki VCF genomu

Pliki VCF genomu (gVCF) to pliki VCF ver. 4.1, które są zgodne ze zbiorem konwencji dotyczących przedstawienia wszystkich miejsc w genomie w dość kompaktowym formacie. W jednym pliku gVCF (*.genome.vcf.gz) zawarte są wszystkie miejsca w docelowym regionie dla określonej próbki.

W pliku gVCF w pozycjach, które nie przeszły wszystkich filtrów, przedstawione są nierozpoznane nukleotydy. Znacznik genotypu (GT) „./.” wskazuje nierozpoznany nukleotyd.

Więcej informacji na ten temat można znaleźć w witrynie sites.google.com/site/gvcftools/home/about-gvcf.

Plik pokrycia amplikonów

Dla każdego pliku wykazu jest generowany plik pokrycia amplikonów. Litera M z liczbą w nazwie pliku oznacza numer wykazu.

Każdy plik zawiera wiersz nagłówka z identyfikatorami próbek powiązanych z wykazem. Plik zawiera następujące informacje:

- ▶ Identyfikator regionu docelowego podany w wykazie.
- ▶ Głębokość pokrycia odczytów, które przeszły przez filtr.

Dodatkowe pliki wyjściowe

Następujące pliki wyjściowe zawierają dodatkowe informacje lub podsumowania wyników przebiegu i błędy analizy. Pliki te nie są konieczne do oceny wyników analizy, jednak można ich użyć na potrzeby rozwiązywania problemów. Wszystkie pliki znajdują się w folderze Alignment (Dopasowanie), chyba że podano inaczej.

Nazwa pliku	Opis
AnalysisLog.txt	Dziennik przetwarzania zawierający opis każdego etapu występującego podczas analizy folderu bieżącego przebiegu. Ten plik nie zawiera komunikatów o błędach. Znajduje się w folderze Alignment (Dopasowanie).
AnalysisError.txt	Dziennik przetwarzania zawierający listę wszystkich błędów, które wystąpiły podczas analizy. Ten plik będzie pusty, jeśli nie wystąpią błędy. Znajduje się w folderze Alignment (Dopasowanie).
DemultiplexSummaryF1L1#.txt	Zawiera raporty wyników demultipleksowania przedstawione w tabeli, w której każdy wiersz odpowiada jednej płytce, a każda kolumna jednej próbce. Znak # odpowiada pasmu 1, 2, 3 lub 4 komory przepływowej. Znajduje się w folderze Alignment (Dopasowanie).
AmpliconRunStatistics.xml	Zawiera zbiorcze dane statystyczne dotyczące danego przebiegu. Znajduje się w folderze Alignment (Dopasowanie).

Folder analizy

Folder analizy zawiera pliki wygenerowane przez oprogramowanie lokalnego menedżera przebiegu.

Związek między folderem wyjściowym i folderem analizy można podsumować w następujący sposób:

- ▶ Podczas sekwencjonowania oprogramowanie do analizy w czasie rzeczywistym (ang. Real-Time Analysis, RTA) wypełnia folder wyjściowy plikami generowanymi podczas analizy obrazu, rozpoznawania nukleotydów i oceny jakościowej.
- ▶ Oprogramowanie RTA kopiuje pliki do folderu analizy w czasie rzeczywistym. Po przypisaniu wyniku jakościowego do każdego nukleotydu na każdy cykl oprogramowanie RTA zapisuje plik RTAComplete.txt w obu folderach.
- ▶ Gdy plik RTAComplete.txt zostanie zapisany, rozpocznie się analiza.
- ▶ W trakcie analizy lokalny menedżer przebiegu zapisuje pliki wyjściowe w folderze analizy, a następnie kopiuje te pliki z powrotem do folderu wyjściowego.

Foldery dopasowań

Za każdym razem, gdy analiza jest ponownie umieszczana w kolejce, lokalny menedżer przebiegu tworzy folder dopasowań o nazwie **Alignment_N**, gdzie „N” jest kolejną liczbą.

Struktura folderów

Alignment (Dopasowanie) – zawiera pliki dopasowania: *.bam, *.vcf, FASTQ oraz pliki swoiste dla modułu analizy.

Date and Time Stamp (Znacznik daty i godziny) – znacznik data_godzina analizy w formacie RRRRMMDD_GGMMSS

- AnalysisError.txt
- AnalysisLog.txt
- aggregate.report.html
- aggregate.report.pdf
- aggregate.summary.csv
- AmpliconCoverage_M#.tsv
- AmpliconRunStatistics.xml
- Sample1.genome.vcf.gz
- Sample1.coverage.csv
- Sample1.report.pdf
- Sample1.summary.csv
- Sample1.vcf.gz
- Sample1.bam
- FASTQ**
 - Sample1**
 - Sample1_L001_R1_001_fastq.gz
 - Stats**
 - DemuxSummaryF1L1.txt
 - FastqSummaryF1L1.txt

Data (Dane)

Intensities (Intensywności)

BaseCalls (Rozpoznanie nukleotydów)

L001 – zawiera pliki *.bcl.

L001 – zawiera pliki *.locs.

RTA Logs – zawiera pliki dziennika z analizy oprogramowania RTA.

InterOp – zawiera pliki binarne służące do raportowania pomiarów sekwencjonowania.

Logs – zawiera pliki dziennika opisujące etapy przeprowadzane podczas sekwencjonowania.

- RTAComplete.txt
- RunInfo.xml
- RunParameters.xml

Rozpoznawanie nukleotydów i różnicowanie indeksu

Podczas sekwencjonowania próbek w aparacie NextSeq 550Dx następuje rozpoznanie nukleotydu (A, C, G lub T) dla każdego klastra danej płytki lub obszaru obrazowania w komorze przepływowej w danym cyklu. W aparacie NextSeq 550Dx wykorzystywane jest dwukanałowe sekwencjonowanie, które wymaga tylko dwóch obrazów do kodowania danych dla czterech nukleotydów DNA, jednego z kanału czerwonego i jednego z kanału zielonego.

Proces odczytu indeksów rozpoznawania nukleotydów różni się od rozpoznawania nukleotydów podczas innych odczytów.

Odczyty indeksów muszą zaczynać się od co najmniej jednego nukleotydu innego niż G w jednym z dwóch pierwszych cykli. Jeśli odczyt indeksu zaczyna się od dwóch rozpoznań nukleotydów G, intensywność sygnału nie jest generowana. Sygnał musi być obecny w obu pierwszych cyklach, aby zapewnić wydajność demultipleksowania.

Jeśli przy wybieraniu indeksów podczas tworzenia przebiegu nie spełniają one wymogów dotyczących różnicowania, pojawi się ostrzeżenie o małym różnicowaniu. Aby zapobiec pojawieniu się tego ostrzeżenia, należy wybrać sekwencje indeksów zapewniające sygnał w obu kanałach dla każdego cyklu.

- ▶ Kanał czerwony – A lub C
- ▶ Kanał zielony – A lub T

Proces rozpoznawania nukleotydów zapewnia dokładność podczas analizowania próbek o niskiej krotności. Więcej informacji na temat sekwencji indeksów zawiera ulotka dołączona do opakowania zestawu *Dx do niestandardowego sekwencjonowania amplikonów TruSeq (nr dokumentu: 1000000029772)*.

Podczas tworzenia przebiegu w lokalnym menedżerze przebiegu wybiera się liczbę próbek do przetestowania. Sugerowane kombinacje indeksów, które spełniają wymogi różnicowania, są wypełniane automatycznie przez oprogramowanie. Chociaż użycie sugerowanych kombinacji indeksów nie jest wymagane, zaleca się skorzystanie z nich.

Historia wersji

Dokument	Data	Opis zmiany
Nr dokumentu: 1000000030329, wer. 04	Sierpień 2021 r.	Zaktualizowano adres autoryzowanego przedstawiciela w UE.
Nr dokumentu: 1000000030329, wer. 03	Grudzień 2019 r.	Zaktualizowano adres autoryzowanego przedstawiciela w UE. Zaktualizowano adres sponsora australijskiego.
Nr dokumentu: 1000000030329, wer. 02	Styczeń 2019 r.	Dodano informacje dotyczące zestawów odczynników w wersji 2.5.
Nr dokumentu: 1000000030329, wer. 01	Sierpień 2018 r.	Zaktualizowano oznaczenia dotyczące zgodności z przepisami.
Nr dokumentu: 1000000030329, wer. 00	Listopad 2017 r.	Pierwsze wydanie.

Pomoc techniczna

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

Witryna: www.illumina.com
Adres e-mail: techsupport@illumina.com

Numery telefonów do działu pomocy technicznej firmy Illumina

Region	Bezpłatne	Regionalne
Ameryka Północna	+1 800 809 4566	
Australia	+1 800 775 688	
Austria	+43 800006249	+43 19286540
Belgia	+32 80077160	+32 34002973
Chiny	400 066 5835	
Dania	+45 80820183	+45 89871156
Finlandia	+358 800918363	+358 974790110
Francja	+33 805102193	+33 170770446
Hiszpania	+34 911899417	+34 800300143
Holandia	+31 8000222493	+31 207132960
Hongkong	800960230	
Irlandia	+353 1800936608	+353 016950506
Japonia	0800 111 5011	
Niemcy	+49 8001014940	+49 8938035677
Norwegia	+47 800 16836	+47 219 39693
Nowa Zelandia	0800 451 650	
Singapur	+1 800 579 2745	
Szwajcaria	+41 565800000	+41 800200442
Szwecja	+46 850619671	+46 200883979
Tajwan	00806651752	
Wielka Brytania	+44 8000126019	+44 2073057197
Włochy	+39 800985513	+39 236003759
Inne kraje	+44 1799 534000	

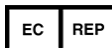
Karty charakterystyki – dostępne na stronie firmy Illumina pod adresem support.illumina.com/sds.html.

Dokumentacja produktu – dostępna do pobrania w formacie PDF w witrynie firmy Illumina. Należy otworzyć stronę support.illumina.com, wybrać produkt, a następnie wybrać **Documentation & Literature** (Dokumentacja i literatura).



Illumina
5200 Illumina Way
San Diego, California 92122, USA
+1 800 809 ILMN (4566)
+1 858 202 4566 (poza Ameryką Północną)
techsupport@illumina.com
www.illumina.com

CE



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

Sponsor australijski

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

DO STOSOWANIA W DIAGNOSTYCE IN VITRO

© 2021 Illumina, Inc. Wszelkie prawa zastrzeżone.

illumina®