

Local Run Manager

Analytický modul Somatic Variant

Průvodce pracovními postupy pro systém NextSeq 550Dx

URČENO K DIAGNOSTICE IN VITRO

Přehled	3
Zadání informací o běhu	4
Analytické metody	6
Zobrazení dat běhu a vzorku	7
Výkaz analýzy	8
Výstupní soubory analýzy	9
Přiřazení báze a diverzita indexů	16
Historie revizí	17
Technická pomoc	18



Tento dokument a jeho obsah je vlastnictvím společnosti Illumina, Inc. a jejích přidružených společností (dále jen „Illumina“). Slouží výlučně zákazníkovi ke smluvním účelům v souvislosti s použitím zde popsaných produktů a k žádnému jinému účelu. Tento dokument a jeho obsah nesmí být používán ani šířen za žádným jiným účelem ani jinak sdělován, zveřejňován či rozmnožován bez předchozího písemného souhlasu společnosti Illumina. Společnost Illumina nepředává tímto dokumentem žádnou licenci na svůj patent, ochrannou známku, autorské právo či práva na základě zvykového práva ani žádná podobná práva třetích stran.

Pokyny v tomto dokumentu musí být důsledně a výslovně dodržovány kvalifikovaným a řádně proškoleným personálem, aby bylo zajištěno správné a bezpečné používání zde popsaných produktů. Veškerý obsah tohoto dokumentu musíte před použitím takových produktů beze zbytku přečíst a pochopit.

NEDODRŽENÍ POŽADAVKU NA PŘEČTENÍ CELÉHO TEXTU A NA DŮSLEDNÉ DODRŽOVÁNÍ ZDE UVEDENÝCH POKYNU MŮŽE VÉST K POŠKOZENÍ PRODUKTŮ, PORANĚNÍ OSOB, AŽ UŽ UŽIVATELŮ ČI JINÝCH OSOB, A POŠKOZENÍ JINÉHO MAJETKU A POVEDE KE ZNEPLATNĚNÍ JAKÉKOLI ZÁRUKY VZTAHUJÍCÍ SE NA PRODUKT.

SPOLEČNOST ILLUMINA NA SEBE NEBERE ŽÁDNOU ODPOVĚDNOST VYPLÝVAJÍCÍ Z NESPRÁVNÉHO POUŽITÍ ZDE POPSANÝCH PRODUKTŮ (VČETNĚ ČÁSTÍ TĚCHTO PRODUKTŮ NEBO SOFTWARE).

© 2021 Illumina, Inc. Všechna práva vyhrazena.

Všechny ochranné známky jsou vlastnictvím společnosti Illumina, Inc. nebo jejích příslušných vlastníků. Informace o konkrétních ochranných známkách naleznete na adrese www.illumina.com/company/legal.html.

Přehled

Modul Somatic Variant softwaru Local Run Manager slouží k použití s rozbořem Illumina TruSeq Custom Amplicon Kit Dx a systémem NextSeq 550Dx. Pokud je rozbor použit s modulem Somatic Variant, slouží k přípravě knihoven používaných k sekvenování DNA z tkáně FFPE (formalinem fixovaná tkáň zalitá v parafinu). Rozbor detekuje somatické mutace v nízkých frekvencích variant.

Analytický modul vyhodnocuje varianty pro krátké oblasti amplifikované DNA, neboli amplikony. Soustředěné sekvenování amplikonů umožňuje vysoké pokrytí konkrétních oblastí ve velkém množství vzorků. Analytický modul provádí sekundární analýzu a generuje výkaz z běhů sekvenování pomocí dvouvláknového přístupu, který zahrnuje dopředné a reverzní oligo fondy. Viz příložená dokumentace sady *TruSeq Custom Amplicon Kit Dx* (dokument č. 1000000029772).

Analytický modul Somatic Variant vyžaduje spotřební materiál pro sekvenování se 300 cykly. Další informace naleznete v příložené dokumentaci sady *NextSeq 550Dx High Output Reagent Kit v2* nebo *NextSeq 550Dx High Output Reagent Kit v2.5*.

Popis této příručky

Tato příručka obsahuje pokyny k nastavení parametrů běhu sekvenování a analýzy pro analytický modul Somatic Variant. Další informace o hlavním panelu softwaru Local Run Manager a nastavení systému naleznete v *Referenční příručce pro přístroj NextSeq 550Dx* (dokument č. 1000000009513).

Zobrazení softwaru Local Run Manager

Rozhraní softwaru Local Run Manager lze zobrazit v operačním softwaru systému NextSeq 550Dx (NOS) nebo ve webovém prohlížeči. Podporovaným webovým prohlížečem je Chromium.



POZNÁMKA

Používáte-li nepodporovaný prohlížeč, objeví se zpráva „Confirm Unsupported Browser“ (Potvrdit nepodporovaný prohlížeč), která vás vyzve ke stažení podporovaného prohlížeče. Vyberte možnost „**here**“ (Zde) a stáhněte si podporovanou verzi Chromia.

Zobrazení na monitoru přístroje

- 1 Chcete-li, aby se rozhraní softwaru Local Run Manager zobrazilo na monitoru přístroje, zvolte některou z následujících možností:
 - ▶ Na domovské obrazovce NOS vyberte program **Local Run Manager**. Po skončení se kliknutím na X v pravém horním rohu vraťte do systému NOS.
 - ▶ Vyberte ikonu Minimize NOS (Minimalizovat NOS), na přístroji otevřete webový prohlížeč Chromium a do adresního řádku zadejte **http://localhost**. Systém NOS mohou minimalizovat pouze uživatelé s oprávněním správce.

Zobrazení z počítače připojeného k síti

- 1 Na počítači s přístupem ke stejné síti, ke které je připojen přístroj, otevřete webový prohlížeč Chromium a pomocí IP adresy nebo názvu přístroje se připojte. Například **http://myinstrument**.

Zadání informací o běhu

Nastavení parametrů

- 1 Přihlaste se k softwaru Local Run Manager.
- 2 Vyberte možnost **Create Run** (Vytvořit běh) a poté možnost **Somatic Variant**.
- 3 Zadejte název běhu, který odlišuje běh od sekvenování prostřednictvím analýzy. Použijte alfanumerické znaky, mezery, podtržítka nebo pomlčky.
- 4 **[Volitelné]** Zadejte popis běhu, který usnadní identifikaci běhu. Použijte alfanumerické znaky, mezery, podtržítka nebo pomlčky.
- 5 V rozevíracím seznamu vyberte počet vzorků a sadu indexů. Při výběru zohledněte následující informace.
 - ▶ Rozevírací seznam obsahuje počty vzorků se sadou indexu. Například 24-Set 1 označuje 24 vzorků k testování s indexy ze sady 1.
 - ▶ Čísla sad indexů označují různé sady indexů i5. Sady Set 1 a Set 2 zajišťují diverzitu indexů. V nabídce jsou dvě sady indexů, což brání vyčerpání jedné sady.
 - ▶ Zvolte počet vzorků, který se nejvíce blíží počtu testovaných vzorků. Pokud není přesný počet vzorků uveden v seznamu, vyberte nejbližší počet, který je nižší než testovaný počet. Pokud například chcete testovat 18 vzorků, vyberte 16 vzorků.
 - ▶ Jamky pro vzorky a kombinace indexů, které splňují požadavky na diverzitu indexů, jsou zvýrazněny zeleně. Pokud při ukládání běhu vyberete jiné jamky a kombinace indexů a požadavky na diverzitu nejsou splněny, zobrazí se upozornění.

Import souborů manifestů pro běh

- 1 Ujistěte se, zda jsou manifesty, které chcete importovat, k dispozici v přístupném síťovém umístění nebo na disku USB.
- 2 Vyberte možnost **Import Manifests** (Importovat manifesty).
- 3 Přejděte k souboru manifestů a vyberte manifesty, které chcete přidat.



POZNÁMKA

Chcete-li zpřístupnit soubory manifestů pro všechny běhy pomocí analytického modulu Somatic Variant, přidejte manifesty pomocí funkce Module Settings (Nastavení modulu). Tato funkce vyžaduje uživatelská oprávnění na úrovni správce. Další informace naleznete v *Referenční příručce pro přístroj NextSeq 550Dx* (dokument č. 1000000009513).


Zadání vzorků pro běh

K zadání vzorků pro běh použijte některou z následujících možností a řiďte se příslušnými pokyny:


- ▶ **Ruční zadání vzorků** – použijte prázdnou tabulku na obrazovce Create Run (Vytvořit běh).
- ▶ **Import vzorků** – vyberte externí soubor ve formátu s hodnotami oddělenými čárkou (*.csv). Na obrazovce Create Run (Vytvořit běh) je k dispozici šablona ke stažení.

Po vyplnění tabulky vzorků můžete exportovat informace o vzorcích do externího souboru. Tento soubor použijte jako referenci při přípravě knihoven nebo importu souboru pro další běh.

Ruční zadání vzorků

- 1 Do pole Sample Name (Název vzorku) zadejte jedinečný název vzorku.
Použijte alfanumerické znaky, pomlčky nebo podtržítka.
Název vzorku se automaticky vyplní v odpovídající jamce druhého fondu.
- 2 **[Volitelné]** Chcete-li nastavit pozitivní nebo negativní kontrolu vzorků, klikněte pravým tlačítkem a vyberte typ kontroly.
Typ kontroly u jamky jednoho vzorku se automaticky vyplní i u odpovídající jamky v druhém fondu.
- 3 **[Volitelné]** Zadejte popis vzorku v poli Sample Description (Popis vzorku).
Použijte alfanumerické znaky, pomlčky nebo podtržítka.
Popis vzorku se automaticky vyplní v odpovídající jamce druhého fondu.
Popisy vzorků jsou přidruženy k ID vzorků. Pokud je v pozdějším běhu znovu použito stejné ID vzorku, jsou popisy vzorků přepsány.
- 4 V rozevíracím seznamu Index 1 (i7) vyberte adaptér Index 1.
Pokud použijete navržené jamky pro vzorky, software automaticky vyplní adaptéry indexů i7 a i5, které splňují požadavky na diverzitu indexů. Pokud není v seznamu uveden přesný počet testovaných vzorků, vyberte adaptéry indexu pro dodatečné jamky. Pokud potřebujete vybrat index pro dodatečné jamky nebo nepoužíváte doporučené kombinace adaptérů indexu, přečtěte si před volbou indexů část *Přiřazení báze a diverzita indexů na straně 16*.
- 5 V rozevíracím seznamu Index 2 (i5) vyberte adaptér Index 2.
- 6 V rozevíracím seznamu Manifest vyberte soubor manifestů.
Vzorky ve fondu A vyžadují jiný manifest než vzorky ve fondu B.
- 7 Zvolte možnost zobrazení, tisku nebo uložení uspořádání desky pro referenci při přípravě knihoven:
 - ▶ Kliknutím na ikonu  **Print** (Tisk) zobrazíte uspořádání desky. Výběrem možnosti **Print** (Tisk) vytisknete uspořádání desky.
 - ▶ Výběrem možnosti **Export** exportujete informace o vzorcích do externího souboru.
Ujistěte se, zda jsou informace o manifestu a vzorcích správné. Nesprávné informace mohou ovlivnit výsledky.
- 8 Vyberte možnost **Save Run** (Uložit běh).

Import vzorků

- 1 Vyberte možnost **Import Samples** (Importovat vzorky) a vyhledejte soubor s informacemi o vzorcích.
Importovat lze dva typy souborů.
 - ▶ Chcete-li vytvořit nové uspořádání desky, vyberte na obrazovce Create Run (Vytvořit běh) možnost **Template** (Šablona). Soubor šablony obsahuje správná záhlaví sloupců pro import. Do každého sloupce pro vzorky běhu zadejte informace o vzorku. Ukázkové informace v nepoužitých buňkách odstraňte a soubor uložte.
 - ▶ Použijte soubor s informacemi o vzorcích, který byl exportován z modulu Somatic Variant pomocí funkce Export.
- 2 Kliknutím na ikonu  **Print** (Tisk) zobrazíte uspořádání desky.
- 3 Výběrem možnosti **Print** (Tisk) vytisknete uspořádání desky pro referenci při přípravě knihoven.

- 4 **[Volitelné]** Výběrem možnosti **Export** exportujete informace o vzorcích do externího souboru. Ujistěte se, zda jsou informace o manifestu a vzorcích správné. Nesprávné informace mohou ovlivnit výsledky.
- 5 Vyberte možnost **Save Run** (Uložit běh).

Úprava běhu

Pokyny k úpravě informací v běhu před sekvenováním naleznete v *Referenční příručce pro přístroj NextSeq 550Dx (dokument č. 100000009513)*.

Analytické metody

Analytický modul Somatic Variant provádí následující kroky analýzy a po jejich dokončení zapisuje výstupní soubory analýzy do složky Alignment (Zarovnání).

- ▶ Demultiplexování čtení indexu
- ▶ Generování souborů FASTQ
- ▶ Zarovnání k referenci
- ▶ Identifikace variant

Demultiplexování

Během demultiplexování se porovnávají jednotlivé sekvence čtení indexu se sekvencemi indexu určenými k běhu. V tomto kroku se neuvažují žádné hodnoty kvality.

Čtení indexu jsou označena v následujících krocích:

- ▶ Vzorky jsou číslovány počínaje hodnotou 1 na základě pořadí, ve kterém jsou uvedeny v běhu.
- ▶ Číslo vzorku 0 je vyhrazeno pro klastry, které nebyly přiřazeny žádnému vzorku.
- ▶ Klastry jsou přiřazovány ke vzorku tehdy, když sekvence indexu přesně souhlasí, nebo v případě nejvýše jednoho nesouladu na čtení indexu.

Generování souborů FASTQ

Po demultiplexování vygeneruje software průběžné soubory analýzy ve formátu FASTQ, které mají textový formát sloužící k vyjádření sekvencí. Soubory FASTQ obsahují čtení pro každý vzorek a přidružená skóre kvality. Klastry, které nespĺnily filtr úspěšnosti, jsou vyloučeny.

Každý soubor FASTQ obsahuje čtení pouze pro jeden vzorek, přičemž název daného vzorku je uveden v názvu souboru FASTQ. Soubory FASTQ jsou hlavním vstupem pro zarovnání. Na každý vzorek a na každý oligo fond se vygeneruje osm souborů FASTQ, čtyři ze čtení 1 a čtyři ze čtení 2. Ve výsledku to znamená celkem 16 souborů FASTQ na vzorek.

Zarovnání

Během kroku zarovnání jsou pomocí Smithova-Watermanova algoritmu zarovnány klastry z každého vzorku vůči sekvencím ampliconů určeným v souboru manifestů.

Smithův-Watermanův algoritmus s pásovou maticí provádí semiglobální zarovnání sekvencí za účelem určení podobných oblastí mezi dvěma sekvencemi. Namísto porovnání celkové sekvence Smithův-Watermanův algoritmus porovnává segmenty všech možných délek.

Každé čtení párového konce je vyhodnoceno z hlediska jeho zarovnání k příslušným zkušebním sekvencím pro dané čtení.

- ▶ Čtení 1 je vyhodnoceno vůči obrácenému doplňku DLSO (Downstream Locus-Specific Oligos).
- ▶ Čtení 2 je vyhodnoceno vůči ULSO (Upstream Locus-Specific Oligos).
- ▶ Pokud začátek čtení souhlasí se zkušební sekvencí s maximálně třemi rozdíly (neshody nebo posuny v důsledku úvodních inzercí/delecí), je celá délka čtení zarovnána vůči ampliconovému cíli pro danou sekvenci.
- ▶ Inzerce/delece v rámci DLSO a ULSO nejsou s ohledem na chemii rozboru pozorovány.

Zarovnání jsou v závislosti na délce ampliconu filtrována z výsledků zarovnání na základě míry neshod buď na oblasti zájmu, nebo na celém ampliconu. Filtrovaná zarovnání jsou zapsána do souborů zarovnání jako nezarovnaná a nejsou použita v přiřazení variant.

Přiřazení variant



Program pro přiřazení variant Pisces Variant Caller, vyvinutý společností Illumina, identifikuje varianty přítomné s nízkou frekvencí ve vzorku DNA.

Program pro přiřazení variant Pisces Variant Caller identifikuje jednonukleotidové varianty (SNV), vícenukleotidové varianty (MNV) a malé inzerce/delece ve třech krocích:

- ▶ Prozkoumá každou pozici v referenčním genomu samostatně.
- ▶ Spočte báze na dané pozici pro zarovnaná čtení, která překrývají pozici.
- ▶ Pomocí Poissonova modelu vypočítá skóre varianty, které určuje kvalitu přiřazení. Varianty se skóre kvality pod Q30 jsou vyloučeny.

Varianty jsou nejdříve přiřazovány pro každý fond odděleně. Poté se varianty z jednotlivých fondů porovnají a zkombinují do jediného výstupního souboru. Pokud je varianta přítomna v obou fondech a splní všechny filtry uvedené v *Poznámky k souborům VCF na straně 12*, označí se varianta v souboru přiřazení variant VCF (Variant Call File) textem PASS (splněno).

Zobrazení dat běhu a vzorku

- 1 Na hlavním panelu softwaru Local Run Manager klikněte na název běhu.
- 2 Na kartě Run Overview (Přehled běhu) si prohlédněte metriky běhu sekvenování.
- 3 **[Volitelné]** Kliknutím na ikonu **Copy to Clipboard**  (Kopírovat do schránky) můžete zkopírovat cestu k výstupní složce běhu.
- 4 Kliknutím na kartu Sequencing Information (Informace o sekvenování) zobrazíte parametry běhu a informace o spotřebním materiálu.
- 5 Kliknutím na kartu Samples and Results (Vzorky a výsledky) zobrazíte umístění výkazu analýzy.
 - ▶ Pokud byla analýza opakována, rozbalte rozevírací seznam Select Analysis (Vyberte analýzu) a vyberte příslušnou analýzu.
- 6 Kliknutím na ikonu **Copy to Clipboard**  (Kopírovat do schránky) můžete zkopírovat cestu ke složce s analýzou.

Další informace o kartách Run Overview (Přehled běhu) a Sequencing Information (Informace o sekvenování) a o postupu opětovného zařazení analýzy naleznete v *Referenční příručce pro přístroj NextSeq 550Dx (dokument č. 1000000009513)*.

Výkaz analýzy

Výsledky analýzy jsou shrnuty na kartě Samples and Results (Vzorky a výsledky) a jako souhrnný výkaz ve složce Alignment (Zarovnání). Pro každý vzorek je také k dispozici soubor výkazu ve formátu PDF.

Informace o kartě Samples and Results (Vzorky a výsledky)

1 Kliknutím na vzorek v seznamu zobrazíte výkaz vzorku.

Tabulka 1 Informace o běhu a vzorku

Záhlaví sloupce	Popis
Run Status (Stav běhu)	Udává, zda byl běh sekvenování úspěšný, nebo neúspěšný.
Total Yield (GB) (Celková výtěžnost (GB))	Počet přiřazení bází v běhu sekvenování. Zobrazuje prahovou hodnotu pro úspěšnost a úspěšný/neúspěšný stav.
% ≥ Q30	Procento čtení v běhu sekvenování se skóre kvality 30(Q30) nebo více. Zobrazuje prahovou hodnotu pro úspěšnost a úspěšný/neúspěšný stav.
Sample Name (Název vzorku)	Název vzorku poskytnutý při vytvoření běhu.
Total PF Reads (Celkový počet čtení PF)	Celkový počet čtení, která splnila filtr úspěšnosti.
Read 1% ≥ Q30 (Čtení 1 – % ≥ Q30)	Procento čtení v běhu 1 se skóre kvality pro daný vzorek 30(Q30) nebo více.
Read 2% ≥ Q30 (Čtení 2 – % ≥ Q30)	Procento čtení v běhu 2 se skóre kvality pro daný vzorek 30(Q30) nebo více.
Autosome Call Rate (Míra přiřazení autozomů)	Počet genomických pozic mezi autozomy (chromozomy 1 až 22), které splňují předdefinovanou prahovou hodnotu spolehlivosti, děleno celkovým počtem vyšetřovaných autozomálních genomických pozic. Míra přiřazení je popisována po jednotlivých vzorcích a vykazována jako procento, které se vypočítá podle vzorce 1 mínus (počet autozomálních pozic s nekompletními přiřazeními děleno celkovým počtem sekvenovaných autozomálních pozic).

Tabulka 2 Informace o výkazu vzorku

Záhlaví sloupce	Popis
Sample (Vzorek)	Název vzorku poskytnutý při vytvoření běhu.
Report Date (Datum výkazu)	Datum, kdy byl výkaz vygenerován.
Informace o vzorku	ID vzorku poskytnuté při vytvoření běhu, celkový počet čtení, která splnila filtr úspěšnosti ve vzorku, procento čtení vzorku se skóre kvality 30(Q30) nebo více a míra přiřazení autozomů.
Amplicon Summary (Souhrn amplikonů)	Celkový počet sekvenovaných oblastí amplikonů, celková délka v bázevých párech sekvenovaných amplikonů v cílových oblastech pro vzorek ve fondu A a fondu B, soubor manifestu použitý pro každý fond. Soubor manifestu určuje referenční genom a cílené referenční oblasti použité v kroku zarovnání.
Read Level Statistics (Statistiky úrovně čtení)	Počet a procento čtení pro vzorek pokrývající každou referenční pozici: pro čtení 1 a čtení 2 ve fondu A a fondu B.
Variants Summary (Souhrn variant)	Počet jednonukleotidových variant, inzercí a delecí detekovaných ve vzorku, které splnily navrhované hodnoty určující, zda výsledky kvality spadají do přijatelného rozsahu.

Záhlaví sloupce	Popis
Coverage Summary (Souhrn pokrytí)	Celkový počet zarovnanýchází děleno velikostí cílené oblasti a procento oblastí ampliconů s hodnotami pokrytí vyššími než je prahová hodnota nízkého pokrytí ($0,2 \times$ průměrné pokrytí ampliconu) pro vzorek ve fondu A a fondu B.
Coverage Plots (Grafy pokrytí)	Graf pokrytí podle oblastí ampliconů znázorňuje pokrytí v oblastech ampliconů daného vzorku. Oblasti, které mají hodnoty pokrytí pod prahovou hodnotou pokrytí, jsou zvýrazněny červeně. Průměr všech hodnot označuje oranžová čára. Graf je poskytnut pro pokrytí fondu A a fondu B.
Software Versions (Verze softwaru)	Verze softwaru při sekvenování vzorku. Zahnuje verze softwaru NextSeq 550Dx Operating Software (NOS), Local Run Manager Software, RTA Software a Somatic Variant Module.

Výstupní soubory analýzy

Pro analytický modul Somatic Variant jsou generovány následující výstupní soubory analýzy. Tyto soubory obsahují výsledky analýzy pro zarovnání a přiřazení variant. Výstupní soubory analýzy jsou umístěny ve složce Alignment (Zarovnání).

Název souboru	Popis
Demultiplexování (*.txt)	Průběžné soubory obsahující souhrnné výsledky demultiplexování.
FASTQ (*.fastq.gz)	Průběžné soubory obsahující přiřazeníází se skóre kvality. Soubory FASTQ jsou hlavním vstupem kroku zarovnání.
Soubory zarovnání ve formátu BAM (*.bam)	Obsahuje zarovnaná čtení pro daný vzorek.
Soubory přiřazení variant Per-pool jsou ve formátu VCF (*.vcf)	Obsahují varianty přiřazené v každé pozici buď z dopředného, nebo reverzního fondu.
Soubory přiřazení variant ve formátu VCF genomu (*.genome.vcf.gz)	Obsahuje genotyp pro každou pozici bez ohledu na to, zda byla přiřazena jako varianta, nebo jako reference.
Soubory přiřazení variant Consensus ve formátu VCF (*.vcf.gz)	Obsahují varianty přiřazené v každé pozici z obou fondů.
AmpliconCoverage_M1.tsv	Obsahuje informace o pokrytí na amplicon a vzorek pro každý poskytnutý manifest. Hodnota M# představuje číslo manifestu.

Formát souboru demultiplexování

Proces demultiplexování čte sekvenci indexu připojenou ke každému klastru, aby určil, ze kterého vzorku klastr pochází. Přiřazení mezi klastry a číslem vzorku je zapsáno do souboru demultiplexování (*.demux) pro každou dlaždici průtokové kyvety.

Formát pojmenování souboru demultiplexování je **s_1_X.demux**, kde X je číslo dlaždice.

Soubory demultiplexování začínají tímto záhlavím:

- ▶ Verze (4bajtové celé číslo), aktuálně 1
- ▶ Počet klastrů (4bajtové celé číslo)

Zbývající část souboru obsahuje čísla vzorků pro jednotlivé klastry ze souboru.

Po dokončení kroku demultiplexování vygeneruje software soubor demultiplexování s názvem **DemultiplexSummaryF1L1.txt**.

- ▶ Text **F1** v názvu označuje číslo průtokové kyvety.
- ▶ Text **L1** v názvu označuje číslo cesty.

- ▶ Výsledky demultiplexování v tabulce s 1 řádkem na každou dlaždici a 1 sloupcem na každý vzorek, včetně vzorku 0.
- ▶ Nejčastěji se vyskytující sekvence ve čteních indexu.

Formát souboru FASTQ

FASTQ je textový formát souboru, který obsahuje přiřazení báze a hodnoty kvality pro jednotlivá čtení. Každý záznam obsahuje 4 řádky:

- ▶ Identifikátor
- ▶ Sekvence
- ▶ Znak +
- ▶ Skóre kvality Phred ve formátu kódování ASCII + 33

Identifikátor má následující formát:

@přístroj:ID_běhu:ID_průtokové_kyvety:cesta:dlaždice:X:Y číslo_čtení:příznak_filtru:0:číslo_vzorku

Příklad:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

Formát souboru BAM

Soubor BAM (*.bam) je komprimovaná binární verze souboru SAM, která slouží k vyjádření zarovnaných sekvencí do velikosti 128 MB. Formáty SAM a BAM jsou podrobně popsány v dokumentu samtools.github.io/hts-specs/SAMv1.pdf.

Soubory BAM využívají formát názvu **NázevVzorku_S#.bam**, kde # je číslo vzorku určené pořadím, ve kterém jsou vzorky uvedeny v běhu.

Soubory BAM obsahují část záhlaví a část zarovnání:

- ▶ **Header** (Záhlaví) – obsahuje informace o celém souboru, jako je název vzorku, délka vzorku a způsob zarovnání. Zarovnání v části zarovnání jsou přidružena ke konkrétním informacím v části záhlaví.
- ▶ **Alignments** (Zarovnání) – obsahuje název čtení, sekvenci čtení, kvalitu čtení, informace o zarovnání a vlastní značky. Název čtení zahrnuje chromozom, počáteční souřadnici, kvalitu zarovnání a řetězec popisovače shody.

Část zarovnání obsahuje následující informace pro každé čtení nebo pár čtení:

- ▶ **AS:** Kvalita zarovnání párového konce
- ▶ **BC:** Značka čárového kódu, která označuje ID demultiplexovaného vzorku přidružené ke čtení.
- ▶ **SM:** Kvalita zarovnání jednoduchého konce.
- ▶ **XC:** Řetězec popisovače shody
- ▶ **XN:** Značka názvu ampliconu, která zaznamenává ID ampliconu přidružené ke čtení

Soubory indexu BAM (*.bam.bai) obsahují index příslušného souboru BAM.

Formát souboru VCF

Formát přiřazení variant (VCF) je běžný formát souboru vyvinutý genomickou vědeckou komunitou. Obsahuje informace o variantách nalezených na konkrétních pozicích v referenčním genomu. Soubory VCF mají příponu .vcf

V záhlaví souboru VCF je uvedena verze formátu souboru VCF a verze programu pro přiřazení variant. Ve zbyvajících částech souboru jsou uvedeny poznámky. Záhlaví souboru VCF také obsahuje soubor referenčního genomu a soubor BAM. Poslední řádek záhlaví obsahuje záhlaví sloupců k datovým řádkům. Každý datový řádek souboru VCF obsahuje informace o jedné variantě.

Záhlaví souboru VCF

Záhlaví	Popis
CHROM (CHROMOZOM)	Chromozom referenčního genomu. Chromozomy se zobrazují ve stejném pořadí jako referenční soubor FASTQ.
POS (POZICE)	Jednobázová pozice varianty v referenčním chromozomu. Pro SNP je tato pozice referenční bází varianty. Pro indely (inzerce/delece) nebo delece je tato pozice referenční bází bezprostředně před variantou.
ID	Číslo rs případné varianty získané ze souboru dbSNP.txt. Pokud je v tomto umístění více čísel rs, jsou položky v seznamu odděleny středníky. Pokud na této pozici neexistuje žádná položka dbSNP, je použita značka chybějící hodnoty (.).
REF (REFERENCE)	Referenční genotyp. Delece jednoho nukleotidu T je například vyjádřena jako referenční TT a alternativní T. Varianta s jedním nukleotidem z A na T je vyjádřena jako referenční A a alternativní T.
ALT	Alely, které se liší od referenčního čtení. Inzerce jednoho nukleotidu T je například vyjádřena jako referenční A a alternativní AT. Varianta s jedním nukleotidem z A na T je vyjádřena jako referenční A a alternativní T.
QUAL (KVALITA)	Skóre kvality podle stupnice Phred přiřazené programem pro přiřazení variant. Vyšší hodnocení značí vyšší spolehlivost varianty a nižší pravděpodobnost chyb. Pro skóre kvality Q je odhadovaná pravděpodobnost chyby $10^{-(Q/10)}$. Například sada přiřazení Q30 má chybovost 0,1 %. Mnoho programů pro přiřazení variant přiřazuje na základě statistických modelů skóre kvality, která jsou vysoká ve vztahu k pozorované chybovosti.

Poznámky k souborům VCF

Záhlaví	Popis
FILTER (FILTR)	<p>Pokud jsou splněny všechny filtry úspěšnosti, zapíše se do sloupce filtru hodnota PASS (splněno).</p> <ul style="list-style-type: none"> • LowDP (Nízká hloubka) – platí u míst, kde je hloubka pokrytí pod 450x v každém fondu. U pozic ampliconů pokrytých dopředným i reverzním čtením se jedná o ekvivalent pokrytí jednoho čtení 900x. • LowGQ (Nízká kvalita genotypu) – kvalita genotypu je pod přípustnou hranicí. • q30 – skóre kvality < 30. • LowVariantFreq (Nízká frekvence variant) – frekvence variant je pod zadanou prahovou hodnotou. • PB – zkreslení fondu sondy (probe pool bias). Varianta nebyla nalezena nebo byla v jednom či dvou fondech sondy nalezena s nízkou frekvencí. • R3x6 – počet sousedících opakování (délky 1 až 3 bp) k přiřazením varianty ≥ 6. • SB – zkreslení prodloužením (strand bias) je větší než zadaná prahová hodnota.
INFO (INFORMACE)	<p>Možné záznamy ve sloupci INFO zahrnují:</p> <ul style="list-style-type: none"> • AC – počet alel (allele count) v genotypech pro každou alternativní alelu ve stejném pořadí, jak jsou uvedeny. • AF – frekvence alel (allele frequency) pro každou alternativní alelu ve stejném pořadí, jak jsou uvedeny. • AN – celkový počet alel v přiřazovaných genotypech. • CD – příznak, který označuje, že se SNP vyskytuje v oblasti kódování alespoň 1 záznamu referenčního genu. • DP – hloubka (počet přiřazení báze zarovnaných k pozici a použitých v přiřazování variant). • Exon – čárkami oddělený seznam oblastí exonů přečtených z referenčního genu. • FC – funkční konsekvence. • GI – čárkami oddělený seznam ID genů přečtených z referenčního genu. • QD – spolehlivost/kvalita variant podle hloubky. • TI – čárkami oddělený seznam ID prepisů přečtených z referenčního genu.
FORMAT (FORMÁT)	<p>Ve sloupci formátu jsou pole oddělená dvojtečkami. Například GT:GQ. Seznam poskytnutých polí závisí na použitém přiřazení variant. Dostupná pole zahrnují:</p> <ul style="list-style-type: none"> • AD – záznam ve tvaru X,Y, kde X je počet referenčních přiřazení a Y je počet alternativních přiřazení. • DP – přibližná hloubka čtení, čtení s MQ=255 nebo špatnými vazbami jsou odfiltrována. • GQ – kvalita genotypu. • GQX – kvalita genotypu. GQX je minimum hodnoty GQ a sloupce QUAL. Obecně platí, že jsou tyto hodnoty podobné. Hodnota GQX přebírá minimum, a je z tohoto důvodu konzervativnějším měřítkem kvality genotypu. • GT – genotyp. 0 odpovídá referenční bázi, 1 odpovídá prvnímu záznamu ve sloupci ALT atd. Lomítko (/) značí, že nejsou k dispozici žádné informace fázování. • NL – úroveň šumu (noise level), odhad šumu v přiřazení bází na této pozici. • PB – zkreslení fondu sondy (probe pool bias). Hodnoty blíží 0 značí větší zkreslení vůči jednomu fondu sondy a menší spolehlivost v přiřazení varianty. • SB – zkreslení prodloužením (strand bias) na této pozici. Větší záporné hodnoty značí menší zkreslení, hodnoty blízko 0 značí větší zkreslení. • VF – frekvence variant (variant frequency), procento čtení podporujících alternativní alelu.
SAMPLE (VZOREK)	Sloupec vzorku udává hodnoty určené ve sloupci formátu.

Soubory VCF genomu

Soubory VCF genomu (gVCF) jsou soubory typu VCF v4.1, které se řídí sadou konvencí pro vyjádření všech míst v rámci genomu v přiměřeně kompaktním formátu. Soubory gVCF (*.genome.vcf.gz) obsahují všechna místa v oblasti zájmu v jediném souboru pro každý vzorek.

Soubor gVCF znázorňuje absence přiřazení na pozicích, které nespĺnily všechny filtry úspěšnosti. Značka genotypu (GT) ./. označuje absenci přiřazení.

Další informace naleznete na webové stránce sites.google.com/site/gvcftools/home/about-gvcf.

Soubory VCF Per-pool a Consensus

Pracovní postup modulu Somatic Variant generuje 2 sady souborů přiřazení variant.

- ▶ **Soubory VCF Per-pool** – obsahují varianty přiřazené buď v dopředném, nebo v reverzním fondu. Soubory Per-pool se zapisují do složky VariantCallingLogs.
- ▶ **Soubory VCF Consensus** – obsahují varianty přiřazené v obou fondech. Soubory Consensus se zapisují do složky Alignment.

Oba typy souborů VCF, Per-pool i Consensus, zahrnují soubory s příponou *.vcf (soubory VCF) i *.genome.vcf (soubory gVCF) a používají následující názvovou konvenci, kde S# představuje pořadí, ve kterém je vzorek připraven pro běh:

- ▶ **Výkazy pro všechna pracoviště** – NavezVzorku_S#.genome.vcf
- ▶ **Výkazy pouze variant** – NavezVzorku_S#.vcf

Software porovnává soubory VCF Per-pool a kombinací dat z každé pozice vytváří soubor VCF Consensus pro vzorek.

Přiřazení variant z obou fondů se do souborů VCF Consensus slučují podle následujících kritérií.

Kritéria	Výsledek
Přiřazení reference v každém fondu	Přiřazení reference
Přiřazení reference v 1 fondu a přiřazení varianty v druhém fondu	Filtrované přiřazení varianty
Odpovídající přiřazení variant s podobnými frekvencemi v obou fondech	Přiřazení varianty
Odpovídající přiřazení variant s významně odlišnými frekvencemi v jednotlivých fondech	Filtrované přiřazení varianty
Neodpovídající přiřazení variant v obou fondech	Filtrované přiřazení varianty

Metriky z obou fondů se slučují pomocí následujících hodnot.

Metrika	Hodnota
Hloubka	Součet hloubek z obou fondů
Frekvence variant	Celkový počet variant děleno celkovou hloubkou pokrytí
Skóre kvality	Minimální hodnota obou fondů

Soubor pokrytí amplikonů

Soubor pokrytí amplikonů je vytvořen pro každý soubor manifestů. Hodnota M# v názvu souboru označuje číslo manifestu.

Každý soubor obsahuje řádek záhlaví, který obsahuje ID vzorků přidružených k manifestu. Soubor obsahuje následující informace.

- ▶ Cílové ID uvedené v manifestu.
- ▶ Hloubka pokrytí čtení, která splnila filtr úspěšnosti.

Doplňující výstupní soubory

Následující výstupní soubory obsahují doplňující informace nebo shrnují výsledky běhu a chyby analýzy. Tyto soubory nejsou nutné k vyhodnocení výsledků analýzy, lze je však použít za účelem řešení problémů. Pokud není uvedeno jinak, jsou všechny soubory umístěny ve složce Alignment (Zarovnání).

Název souboru	Popis
AnalysisLog.txt	Protokol zpracování, který popisuje jednotlivé kroky analýzy aktuální složky běhu. Tento soubor neobsahuje chybové zprávy. Nachází se ve složce Alignment (Zarovnání).
AnalysisError.txt	Protokol zpracování, který obsahuje případné chyby analýzy. Pokud nedošlo k žádným chybám, bude tento soubor prázdný. Nachází se ve složce Alignment (Zarovnání).
DemultiplexSummaryF1L1#.txt	Výkaz výsledků demultiplexování v tabulce s 1 řádkem na každou dlaždici a 1 sloupcem na každý vzorek. Znak # je nahrazen číslem cesty 1, 2, 3 nebo 4 průtokové kvety. Nachází se ve složce Alignment (Zarovnání).
AmpliconRunStatistics.xml	Obsahuje souhrnnou statistiku týkající se běhu. Nachází se ve složce Alignment (Zarovnání).

Složka Analysis (Analýza)

Ve složce Analysis (Analýza) se nacházejí soubory vytvořené softwarem Local Run Manager.

Vztah mezi výstupní složkou a složkou analýzy lze shrnout takto:

- ▶ Během sekvenování vyplní proces analýzy v reálném čase (RTA) výstupní složku soubory, které byly vygenerovány během analýzy snímků, přiřazení báze a vyhodnocování kvality.
- ▶ Proces RTA zkopíruje soubory do složky analýzy v reálném čase. Jakmile proces RTA přiřadí skóre kvality každé bázi pro každý cyklus, software zapíše soubor RTAComplete.txt do obou složek.
- ▶ Pokud je přítomen soubor RTAComplete.txt, bude zahájena analýza.
- ▶ Během pokračující analýzy zapíše software Local Run Manager výstupní soubory do složky analýzy a poté zkopíruje soubory zpět do výstupní složky.

Složky Alignment (Zarovnání)

Při každém opětovném zařazení analýzy vytvoří software Local Run Manager složku Alignment (Zarovnání) s názvem **Alignment_N**, kde N je pořadové číslo.

Struktura složek

Alignment (Zarovnání) – obsahuje soubory *.bam, *.vcf, FASTQ a soubory týkající se analytického modulu.

Date and Time Stamp (Označení data a času) – Označení data a času analýzy ve formátu RRRRMMDD_HHMMSS

- 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 (Vzorek1)

- Sample1_L001_R1_001_fastq.gz

Stats (Statistika)

- DemuxSummaryF1L1.txt
- FastqSummaryF1L1.txt

Data

Intensities (Intenzity)

BaseCalls (Přirazení báze)

L001 – obsahuje soubory *.bcl.

L001 – obsahuje soubory *.locs.

RTA Logs (Protokoly RTA) – obsahuje soubory protokolů z analýzy softwaru RTA.

InterOp – obsahuje binární soubory sloužící k vykazování metrik běhů sekvenování.

Logs (Protokoly) – obsahuje soubory protokolu popisující kroky provedené během sekvenování.

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

Přiřazení báze a diverzita indexů

Při sekvenování vzorků přístrojem NextSeq 550Dx přiřazení bází určuje bázi (A, C, G nebo T) pro každý klastr dané dlaždice nebo snímací oblast průtokové kyvety, a to pro konkrétní cyklus. Přístroj NextSeq 550Dx používá dvoukanálové sekvenování, které vyžaduje k zakódování dat pro čtyři báze DNA pouze dva snímky, jeden z červeného a jeden ze zeleného kanálu.

Proces čtení indexu přiřazení báze se liší od přiřazení báze během jiných čtení.

Čtení indexu musí začínat nejméně jednou bází jinou než G ve kterémkoliv z prvních dvou cyklů. Pokud čtení indexu začíná dvěma přiřazeními báze G, není generována žádná intenzita signálu. Signál musí být přítomen v každém z prvních dvou cyklů, aby bylo zajištěno řádné provedení demultiplexování.

Pokud při výběru indexů během vytváření běhu nesplní indexy požadavky na diverzitu, zobrazí se upozornění na nízkou diverzitu. Chcete-li zabránit zobrazení upozornění na nízkou diverzitu, vyberte sekvence indexu, které poskytují signál na obou kanálech pro každý cyklus.

- ▶ Červený kanál – A nebo C
- ▶ Zelený kanál – A nebo T

Tento proces přiřazení bází zajišťuje přesnost při analýze malého počtu vzorků. Další informace o sekvencích indexů naleznete v příložené dokumentaci sady *TruSeq Custom Amplicon Kit Dx* (dokument č. 1000000029772).

Během vytváření běhu v softwaru Local Run Manager zvolíte počet vzorků k testování. Navržené kombinace indexů, které splňují požadavky na diverzitu indexu, jsou softwarem vyplněny automaticky. Ačkoliv použití navržených kombinací indexů není požadováno, je doporučeno.

Historie revizí

Dokument	Datum	Popis změny
Dokument č. 1000000030330 v04	srpen 2021	Aktualizována adresa oprávněného zástupce v EU.
Dokument č. 1000000030330 v03	Duben 2020	Aktualizována adresa oprávněného zástupce v EU. Aktualizována adresa australského sponzora.
Dokument č. 1000000030330 v02	Leden 2019	Byly doplněny informace o sadách reagentů v2.5.
Dokument č. 1000000030330 v01	srpen 2018	Byla aktualizována regulační označení.
Dokument č. 1000000030330 v00	listopad 2017	První vydání.

Technická pomoc

Pokud potřebujete technickou pomoc, obraťte se na technickou podporu společnosti Illumina.

Web: www.illumina.com
E-mail: techsupport@illumina.com

Telefonní čísla na zákaznickou podporu společnosti Illumina

Oblast	Bezplatná linka	Regionální linka
Severní Amerika	+1 800 809 4566	
Austrálie	+1.800.775.688	
Belgie	+32 800 771 60	+32 340 029 73
Čína	400 066 5835	
Dánsko	+45 808 201 83	+45 898 711 56
Finsko	+358 800 918 363	+358 974 790 110
Francie	+33 805 102 193	+33 170 770 446
Hongkong	800960230	
Irsko	+353 180 093 6608	+353 016 950 506
Itálie	+39 800 985 513	+39 236 003 759
Japonsko	0800.111.5011	
Německo	+49 800 101 4940	+49 893 803 5677
Nizozemsko	+31 800 022 2493	+31 207 132 960
Norsko	+47 800 168 36	+47 219 396 93
Nový Zéland	0800.451.650	
Rakousko	+43 800 006 249	+43 192 865 40
Singapur	+1 800 579 2745	
Španělsko	+34 911 899 417	+34 800 300 143
Spojené království	+44 800 012 6019	+44 207 305 7197
Švédsko	+46 850 619 671	+46 200 883 979
Švýcarsko	+41 565 800 000	+41 800 200 442
Tchaj-wan	00806651752	
Ostatní země	+44.1799.534000	

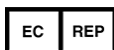
Bezpečnostní listy (SDS) – k dispozici na webu společnosti Illumina na adrese support.illumina.com/sds.html.

Dokumentace k produktu – je k dispozici ke stažení z webu společnosti Illumina ve formátu PDF. Přejděte na web support.illumina.com, vyberte produkt a potom vyberte možnost **Documentation & Literature** (Dokumentace a literatura).



Illumina
5200 Illumina Way
San Diego, Kalifornie 92122 U.S.A.
+1 800 809 ILMN (4566)
+1 858 202 4566 (mimo Severní Ameriku)
techsupport@illumina.com
www.illumina.com

CE



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

Australský sponzor

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

URČENO K DIAGNOSTICE IN VITRO

© 2021 Illumina, Inc. Všechna práva vyhrazena.

illumina®