

illumina®

TruSight Whole Genome Analysis Application

Dokumentace k produktu

VLASTNICTVÍ SPOLEČNOSTI ILLUMINA

Dokument č. 200049931 v00

Duben 2024

URČENO K DIAGNOSTICE IN VITRO.

Historie revizí

| Dokument | Datum | Popis změny |
|---------------------------|---------------|---------------|
| Dokument č. 200049931 v00 | Duben 2024 | První vydání. |

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 kterýchkoli 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 POKYŇŮ 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 DÍLŮ TĚCHTO PRODUKTŮ NEBO SOFTWARE).

© 2024 Illumina, Inc. Všechna práva vyhrazena.

Všechny ochranné známky jsou vlastnictvím společnosti Illumina, Inc. nebo jejích příslušných vlastníků. Podrobné informace o ochranných známkách viz adresa www.illumina.com/company/legal.html.

Obsah

| | |
|--|-----------|
| Historie revizí | ii |
| Přehled | 1 |
| Začínáme | 1 |
| Nastavení | 2 |
| Vytvoření běhu | 4 |
| Revize běhu | 5 |
| Opětovné zařazení analýzy | 6 |
| Zobrazení běhu a výsledků | 7 |
| Souhrn výstupního souboru | 8 |
| Informace o výkazu kontroly kvality | 8 |
| Informace o souboru přiřazení varianty | 12 |
| Soubory FASTQ | 20 |
| Soubory CRAM | 21 |
| Technická pomoc | 22 |

Přehled

TruSight Whole Genome Analysis Application se používá k plánování sekvenačních analýz pro TruSight Whole Genome a k automatickému zahájení analýzy po dokončení cyklu. Analýza zahrnuje demultiplexování, generování FASTQ, mapování čtení, zarovnání s lidským referenčním genomem GrCh38/hg38 s podporou grafu a přiřazení variant pomocí systému Illumina DRAGEN Server for NovaSeq 6000Dx.

V různých fázích pracovního postupu analýzy provádí aplikace kontrolu kvality (QC) podle definovaného sekvenování, FASTQ a metrik knihovny vzorků a generuje výkazy s výsledky. U vzorků, které projdou všemi kroky kontroly kvality, aplikace generuje podpůrné výstupní soubory pro použití v následných aplikacích analýzy germinálního genomu.

Aplikace TruSight Whole Genome Analysis Application spouští detektory variant systému DRAGEN, včetně detektoru malých variant, detektoru variant počtu kopií (CNV) a detekce expanze repetice pomocí systému ExpansionHunter.

Aplikace také provádí anotaci nízké, střední nebo vysoké úrovně spolehlivosti pro malé varianty a tuto anotaci zahrne do výstupního souboru.

Začínáme

Ujistěte se, že aplikace TruSight Whole Genome Analysis Application je nainstalována v NovaSeq 6000Dx instrument, který bude použit pro sekvenování, jako součást systému TruSight Whole Genome. Nainstalované aplikace naleznete na obrazovce Applications (Aplikace) v přístroji NovaSeq 6000Dx Instrument nebo v systému Illumina Run Manager pomocí prohlížeče na počítači připojeném k síti. Potřebujete-li pomoc s naplánováním instalace, obraťte se na místního zástupce společnosti Illumina.

Požadavky na ukládání dat

Obecné informace o výstupu a ukládání dat naleznete v Dokumentace k přístroji NovaSeq 6000Dx (dokument č. 200010105) a Dokumentace k produktu DRAGEN Server for NovaSeq 6000Dx (dokument č. 200014171).

Aplikace TruSight Whole Genome Analysis Application ukládá data do složky Run (Běh) a do složky Analysis (Analýza) v externím úložišti. Minimální požadavky na úložiště lze přibližně odhadnout z velikosti výstupu dat do každé složky pro jeden sekvenační běh uvedený níže.

| Konfigurace | Složka běhu (GB) | Složka analýzy (GB) |
|---------------------------------|------------------|---------------------|
| Průtoková kyveta S2 (6 vzorků) | přibl. 430 | přibl. 350 |
| Průtoková kyveta S4 (16 vzorků) | přibl. 1 110 | přibl. 890 |

Přibližná doba analýzy

Analýza se spustí automaticky po dokončení sekvenačního cyklu a probíhá postupně u vzorků v rámci běhu. Výstupní soubory dat budou k dispozici v externím úložišti po dokončení analýzy pro všechny vzorky v běhu a po dokončení přenosu kopie do externího úložiště. Při spuštění sekvenačního běhu na obou stranách A a B současně bude sekvenování probíhat souběžně. Analýzu těchto sekvenačních běhů provede aplikace TruSight Whole Genome Analysis Application postupně po dokončení sekvenování. Běh, jehož sekvenování a přenos se dokončí jako první, bude analyzován jako první. Druhý běh sekvenování bude přenesen a zařazen do fronty pro analýzu po dokončení první analýzy. Informace o určení stavu aktivních nebo neúspěšných běhů naleznete v části [Zobrazení běhu a výsledků na straně 7](#).

Přibližná doba, než budou k dispozici výsledky analýzy po dokončení sekvenování, je uvedena níže pro situaci, kdy jsou strana A i strana B naplněny současně se stejnou konfigurací.

| Konfigurace | Analýza 1. běhu | Analýza 2. běhu |
|------------------------------------|-----------------|-----------------|
| Průtoková kyveta S2 (6 vzorků) | přibl. 12 hodin | přibl. 24 hodin |
| Průtoková kyveta S4 (16 vzorků) | přibl. 24 hodin | přibl. 48 hodin |

Nastavení

Pro zobrazení aktuální konfigurace a změnu oprávnění vyberte TruSight Whole Genome Analysis Application na obrazovce Applications (Aplikace).

Konfigurace

Konfigurační obrazovka zobrazuje následující nastavení aplikace:

- **Application Name** (Název aplikace)
- **Application Version** (Verze aplikace)
- **DRAGEN Version** (Verze)
- **RTA Version** (Verze RTA)
- **Release Date** (Datum vydání)
- **Organization** (Organizace)
- **Device Identifier** (Identifikátor prostředku)
- **Production Identifier** (Identifikátor výroby)
- **Library Prep Kits** (Sady pro přípravu knihoven) – zobrazuje sadu pro přípravu knihovny. Toto nastavení nelze změnit.

- **Index Adapter Kits** (Sady indexových adaptérů) – zobrazuje sady indexových adaptérů, které lze použít.
- **Index Reads** (Čtení indexů)
- **Read Type** (Typ čtení)
- **Index Lengths** (Délky indexů)
- **Read Lengths** (Délky čtení) – délky čtení jsou nastaveny ve výchozím nastavení při výběru sady indexů. Toto nastavení nelze změnit.

Oprávnění

Určený správce má přístup k položce Permissions (Oprávnění) a může pomocí zaškrtačacích políček na obrazovce Permissions (Oprávnění) spravovat přístup uživatelů k aplikaci TruSight Whole Genome Analysis Application.

Další informace o oprávněních a správě uživatelů naleznete v části Konfigurace systému v Dokumentace k přístroji NovaSeq 6000Dx (dokument č. 200010105).

Vytvoření běhu

Nové běhy v režimu IVD můžete vytvářet buď na přístroji, nebo přístupem k Illumina Run Manager (IRM) pomocí prohlížeče na počítači připojeném k síti. Pro vzdálený přístup k přístroji použijte adresu a informace o uživatelském účtu poskytnuté zástupcem společnosti Illumina. Další informace viz Dokumentace k přístroji NovaSeq 6000Dx (dokument č. 200010105).

Create Run (Vytvořit běh) je doporučená metoda pro plánování běhů. Použití možnosti Import Sample Sheet (Importovat seznam vzorků) se nedoporučuje. Soubory seznamů vzorků ve složkách běhu a analýzy nejsou vhodné pro import během plánování běhu.

Vytvoření běhů

1. Na obrazovce Runs (Běhy) vyberte možnost **Create Run** (Vytvořit běh).
2. Vyberte aplikaci TruSight Whole Genome Analysis Application a potom vyberte **Next** (Další).
3. Na obrazovce Run Settings (Nastavení běhu) zadejte název běhu. Název běhu identifikuje běh od sekvenování až po analýzu.
4. [Volitelné] Pro další identifikaci běhu zadejte popis běhu. Souprava pro přípravu knihovny je ve výchozím nastavení nastavena jako TruSight Whole Genome a nelze ji změnit.
5. Vyberte požadovanou sadu indexů systému TruSight Whole Genome z rozevírací nabídky **Index Adapter Kit** (Sada indexového adaptéru).
Délka čtení bude nastavena podle výchozího nastavení a nelze ji změnit. (Čtení 1 a 2 používá 151 cyklů; indexování 1 a 2 používá 10 cyklů).
6. Zadejte ID knihovny zkumavek (doporučený formát jako DX1234567-LIB) a poté vyberte **Next** (Další).
Pokud v tomto kroku není specifikováno žádné ID knihovny zkumavek, bude nutné před založením sekvenčního spotřebního materiálu vybrat plánovaný běh. Pokud je v tomto kroku zadáno nesprávné ID knihovny zkumavek, musí se plánovaný běh před vložením spotřebního materiálu opravit. Viz [Revize běhu na straně 5](#), kde je uveden protokol pro opravu běhu, když jste připraveni vložit spotřební materiál.
7. Na obrazovkách Sample Data (Data vzorku) a Sample Settings (Nastavení vzorku) budou zadány informace o vzorku. Data vzorků lze zadat ručně nebo importem souboru seznamu vzorků. ID vzorku musí být pro každý vzorek jedinečné a může obsahovat pouze alfanumerické znaky, podtržítka a pomlčky. Nepoužívejte mezery. Well Position (Pozice jamky) odkazuje na jamku ve formátu A01 až H04 indexovací destičky. Po zadání pozice jamky indexovací destičky se automaticky vyplní informace o sekvenci indexů. Pohlaví musí být zadáno jako Male (Muž), Female (Žena) nebo Unknown (Není známo). Library Plate ID (ID destičky knihovny) a Library Well ID (ID jamky knihovny) (např. formát A01) jsou povinná pole.

- Chcete-li zadat data vzorku ručně, přidejte řádky (celkem 6 pro kyvetu S2 nebo 16 pro kyvetu S4) a zadejte požadované informace do polí Sample ID (ID vzorku) a Well Position (Poloha jamky). Informace lze také kopírovat a vkládat z aplikace Excel. Vyberte možnost **Next** (Další). Na obrazovce Sample Settings (Nastavení vzorku) zadejte Library Plate ID (ID destičky knihovny), Library Well ID (ID jamky knihovny) a Sex (Pohlaví). Vyberte možnost **Next** (Další).
 - Chcete-li importovat soubor s údaji vzorků, vyberte možnost **Import Samples** (Importovat vzorky) a nahrajte soubor s údaji vzorků. Informace budou automaticky vyplněny do řádků. Na této obrazovce je k dispozici ke stažení šablona (*.csv). Vyberte možnost **Next** (Další). Na obrazovce Sample Settings (Nastavení vzorku) budou informace automaticky vyplněny do řádků z importovaného souboru s údaji vzorků. Vyberte možnost **Next** (Další).
8. Na obrazovce Analysis Settings (Nastavení analýzy) zadejte název dávky zaznamenaný během plánování dávky a běhu.
 9. [Volitelné] Vyberte typ průtokové kyvety, S2 nebo S4.
 10. Potvrďte nebo zrušte zaškrtnutí políčka pro generování souborů FASTQ komprimovaných metodou ORA a poté vyberte **Next** (Další).

POZNÁMKA Ve výchozím nastavení TruSight Whole Genome Analysis Application generuje soubory FASTQ komprimované metodou ORA. Změna tohoto nastavení zvětší velikost konečného datového výstupu.

11. Na obrazovce Run Review (Kontrola běhu) zkontrolujte zadané informace. Pokud nejsou nutné žádné změny, vyberte možnost **Save** (Uložit). Pokud jsou nutné změny, vyberte podle potřeby **Back** (Zpět) a vraťte se na příslušnou obrazovku.



UPOZORNĚNÍ

TruSight Whole Genome byl validován pro 6 vzorků při použití průtokové kyvety NovaSeq 6000Dx S2 a 16 vzorků při použití průtokové kyvety NovaSeq 6000Dx S4. Ujistěte se, že je zadán správný počet vzorků pro vybranou konfiguraci průtokové kyvety.

Revize běhu

Pokud jsou po vytvoření běhu a před vložením spotřebního materiálu pro sekvenování nutné změny, revidujte běhy v režimu IVD buď na přístroji, nebo přístupem k Illumina Run Manager (IRM) pomocí prohlížeče na síťovém počítači.

1. Vyberte možnost **Runs** (Běhy).
2. Vyberte název běhu na kartě Planned Runs (Plánované běhy).
3. Vyberte možnost **Edit** (Upravit).
4. Podle potřeby aktualizujte informace o běhu nebo vzorku. Například zadejte nebo opravte Library Tube ID (ID zkumavky knihovny) tak, aby odpovídalo tomu, které bylo použito při přípravě pracovního postupu.

5. Vybírejte možnost **Next** (Další) až po Run Review (Revize běhu).
6. Vybete možnost **Save** (Uložit).
7. Vybete možnost **Exit** (Ukončit).

Vraťte se k sekvenování v režimu IVD a opakujte vložení spotřebního materiálu. Běh by měl být nyní automaticky zvýrazněn.

Pokud při vkládání spotřebního materiálu aktualizujete Library Tube ID (ID zkumavky knihovny), vraťte se k možnosti Run Selection (Výběr běhu) v Control Software (Řídicí software) a vyberte možnost **Refresh** (Obnovit) pro související sloupec A nebo B. Běh by měl být nyní automaticky zvýrazněn. Pokud není, vyberte **Back** (Zpět) a zopakujte vložení spotřebního materiálu.

Opětovné zařazení analýzy

Informace o tom, který typ opětovného zařazení analýzy je nejvhodnější, naleznete v části Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132).

Opětovné zařazení analýzy beze změn

1. Vybete možnost Completed Runs (Dokončené běhy), aby se zobrazily Run Details (Podrobnosti běhu).
2. Vybete možnost **Requeue Analysis** (Znovu zařadit analýzu).
3. Vybete možnost **Requeue Analysis with no changes** (Znovu zařadit analýzu beze změn).
4. Zadejte podrobnosti do pole Reanalysis Reason (Důvod opětovné analýzy).
5. Vybete možnost **Requeue Analysis** (Znovu zařadit analýzu).
6. Opusťte stávající stránku a přejděte na stránku Active Runs (Aktivní běhy), abyste se ujistili, že opětovné zařazení analýzy probíhá.

Opětovné zařazení analýzy se změnami

1. Vybete možnost Completed Runs (Dokončené běhy), aby se zobrazily Run Details (Podrobnosti běhu).
2. Vybete možnost **Requeue Analysis** (Znovu zařadit analýzu).
3. Vybete možnost **Edit run settings** (Upravit nastavení běh) a **Requeue Analysis** (Znovu zařadit analýzu).
4. Zadejte podrobnosti do pole Reanalysis Reason (Důvod opětovné analýzy).
5. Vybete možnost **Requeue Analysis** (Znovu zařadit analýzu).
6. Potvrďte nebo aktualizujte Run Settings (Nastavení běhu) a poté vyberte možnost **Next** (Další).

7. Opravte informace o vzorku podle potřeby ruční aktualizací polí nebo vyberte možnost **Download Template** (Stáhnout šablonu) a vytvořte soubor `sampledata.csv` s aktuálními informacemi. Opravte informace a odstraňte stávající řádky na kartě Sample Data (Data vzorku) předtím, než použijete možnost Import Samples (Importovat vzorky) k vyplnění opravených dat vzorku.
8. Na obrazovce Run Review (Kontrola běhu) zkontrolujte informace o běhu a výběrem možnosti **Save** (Uložit) spusťte opakovanou analýzu.
9. Vyberte možnost **Exit** (Ukončit) a přejděte na stránku Active Runs (Aktivní běhy), abyste se ujistili, že opětovné zařazení analýzy probíhá.
Aby mohla být opakovaná analýza úspěšně dokončena, musí být původní složka dat běhu přítomna v externím úložišti specifikovaném v části Run Details (Podrobnosti běhu). Pokud se opětovná analýza nezdaří, ujistěte se, že běh nebyl přesunut ani odstraněn.

Zobrazení běhu a výsledků

1. Na hlavní obrazovce Illumina Run Manager v režimu IVD vyberte možnost **Runs** (Běhy).
2. Na kartě Completed Runs (Dokončené běhy) vyberte název běhu.
Na této kartě se také zobrazí běhy, které byly ukončeny z důvodu selhání sekvenování, přenosu dat nebo analýzy. Aktivní běhy a jejich stav se zobrazují na kartě Active Runs (Aktivní běhy). Další informace viz Dokumentace k přístroji NovaSeq 6000Dx (dokument č. 200010105).
3. Výběrem názvu běhu na kartě Completed Runs (Dokončené běhy) zobrazíte Run Details (Podrobnosti běhu) a Results (Výsledky) pro cestu k výstupní složce analýzy.
U neúspěšných běhů zkontrolujte Status (Stav) každého kroku a poté se podívejte do části Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132).
4. Přejděte do složky analýzy na místním disku a otevřete Consolidated Report (Konsolidovaná zpráva), abyste si prohlédli výsledek PASS/FAIL (ÚSPĚŠNÝ/NEÚSPĚŠNÝ) pro každý krok kontroly kvality následujícím způsobem:
 - Informace o kontrole kvality sekvenčního běhu naleznete v Summary Sequencing QC Result (Souhrnný výsledek kontroly kvality sekvenování).
 - Informace o kontrole kvality FASTQ pro každý vzorek v běhu naleznete v Summary FASTQ QC Result (Souhrnný výsledek kontroly kvality FASTQ).
 - Informace o kontrole kvality knihovny pro každý vzorek v běhu naleznete v Summary Sample Library QC Result (Souhrnný výsledek kontroly kvality knihovny).Pokud je zjištěn výsledek FAIL (NEÚSPĚŠNÝ), poznamenejte si krok kontroly kvality a přejděte k části Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132).

Souhrn výstupního souboru

Aplikace TruSight Whole Genome Analysis Application uloží následující hlavní výstupní soubory. Umístění hlavních výstupních souborů naleznete v částech s informacemi o souborech níže.

Pro běhy a vzorky, které nesplňují kritéria platnosti, se nevytvářejí soubory CRAM, ROH ani *genome.vcf.

| Výstupní soubor | Popis |
|--|--|
| Konsolidovaný výkaz (*.csv) | Obsahuje metriky kvality používané pro validitu běhu (včetně celkového výtěžku a Q30), metriky validity vzorku (včetně výtěžku FASTQ), metriky kontroly kvality knihovny a metriky For Information Only (FIO) (Pouze pro informaci) pro všechny vzorky v běhu. |
| Vzorový výkaz (*.csv) | Obsahuje výsledky sekvenční QC, FASTQ a QC knihovny vzorků. Výkaz také obsahuje konkordanci ploidie a metriky FIO pro jednotlivé vzorky a také související sekvenační běh. |
| Malá varianta a mSNV VCF (*.annotated.hard-filtered-gvcf.gz) | Obsahuje informace o přiřazení pro malé varianty (SNV, indels) a mitochondriální SNV. |
| CNV VCF (*.cnv.vcf.gz) | Obsahuje informace o přiřazení varianty pro zisky a ztráty počtu kopií. |
| Repeats VCF (*.repeats.vcf.gz) | Obsahuje informace o přiřazení variant na rozšířeních STR a SMN1. |
| ROH BED (*.roh.bed) | Obsahuje informace pro oblasti homozygoty. |
| FASTQ (*.fastq.gz nebo *.fastq.ora) | Přechodné soubory obsahující přiřazení bází se skórem kvality. Soubory FASTQ jsou primárním vstupem pro krok zarovnání. Pokud je vybrána komprese ORA, odrazí se to v názvu souboru. |
| CRAM zarovnání (*.cram) | Obsahuje zarovnaná čtení pro daný vzorek. |

Informace o výkazu kontroly kvality

Konsolidovaný výkaz `<<RunID>>_Consolidated_Report.csv` se nachází v adresáři `TruSightWholeGenomeAnalysis_x.x.x_run-complete` a obsahuje informace o metrikách kvality používaných k určení, zda vzorky splňují nebo nesplňují požadavky v různých fázích analýzy. Výkazy jednotlivých vzorků `<<Sample_ID>>_Sample_Report.csv` naleznete ve složkách `<Sample_ID>` v adresáři `TruSightWholeGenomeAnalysis_x.x.x_run-complete`.

Záhlaví výkazu obsahují následující informace o běhu: verze aplikace, název dávky, ID zkumavky směsi knihovny, název sekvenčního běhu, ID sekvenčního běhu a typ průtokové kyvety. Následující tabulky popisují informace obsažené v konsolidovaném výkazu. Jednotlivý vzorový výkaz obsahuje stejné informace s výjimkou demultiplexní metriky.

Tabulka 1 Metriky kontroly kvality sekvenování

| Metrika | Specifikace | Popis |
|--|--------------------------------------|---|
| Neindexovaná celková výtěžnost (GB) | — | Žádná specifikace, protože běhy s nižší výtěžností mohou vést k tomu, že knihovny vzorků budou vyhovující. Očekávejte $\geq 3\,000$ Gbp pro S4 a $\geq 1\,000$ GB pro průtokovou kyvetu S2. |
| Celkové % $\geq Q30$ | ≥ 85 | Míra kvality bází na úrovni běhu. Je stanovena minimální specifikace, protože příliš nízké %Q30 běhů bude mít za následek, že báze Q30 neprojdou při kontrole kvality knihovny vzorků. |
| Souhrnný výsledek kontroly kvality sekvenování | PASS (Úspěšné) nebo FAIL (Neúspěšné) | V případě nesplnění požadavků kontroly kvality si prostudujte část Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132). |

Tabulka 2 Metriky demultiplexování

| Metrika | Specifikace | Popis |
|---------------------------------|-------------|---|
| Procento identifikovaných čtení | — | Celkový zlomek čtení v běhu, které prošly filtrem, jež byly přiřazeny ke vzorkům během demultiplexování. |
| Procento CV | — | Poskytuje měřítko rovnoměrnosti demultiplexovaných čtení pro každý indexový pár v běhu. U cyklů bez selhání výsledku kontroly kvality FASTQ očekávejte hodnotu $< 25\%$. |

Tabulka 3 Metriky kontroly kvality FASTQ

| Metrika | Specifikace | Popis |
|--|--------------------------------------|---|
| Výtěžnost na vzorek (bps) | $\geq 90\,000\,000\,000$ | Minimální hodnota je nastavena tak, aby odpovídala přibližně 26násobnému průměrnému autozomálnímu pokrytí za účelem třídění knihoven vzorků, které neprojdou kontrolou kvality, aby se zkrátila doba analýzy. |
| Souhrnný výsledek kontroly kvality FASTQ | PASS (Úspěšné) nebo FAIL (Neúspěšné) | V případě nesplnění požadavků kontroly kvality FASTQ si prostudujte část Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132). |

Tabulka 4 Metriky kontroly kvality knihovny vzorků

| Metrika | Specifikace | Popis |
|--|--------------------------------------|---|
| Průměrné autozomální pokrytí | ≥ 35 | Průměrné pokrytí v rámci všech autozomů. Minimální specifikace je nastavena tak, aby byla zajištěna analytická účinnost. |
| Procento autozomů s pokrytím větším než 20X | $\geq 93,94$ | Míra rovnoměrnosti pokrytí, které detekuje problémy, které nemusí nutně souviset se zkreslením GC. Minimální specifikace je nastavena tak, aby byla zajištěna analytická účinnost. |
| Normalizované pokrytí při 60 % až 79 % dílů GC | $0,82 \leq x \leq 1,13$ | Míra rovnoměrnosti pokrytí, která detekuje zkreslení GC, konkrétně ztrátu pokrytí v oblastech genomu s vyšším procentuálním podílem GC bází a nižším procentuálním podílem AT bází. Minimální a maximální specifikace je nastavena tak, aby byla zajištěna analytická účinnost. |
| Normalizované pokrytí při 20 % až 39 % dílů GC | $0,97 \leq x \leq 1,06$ | Míra rovnoměrnosti pokrytí, která detekuje zkreslení GC, konkrétně ztrátu pokrytí v oblastech genomu s nižším procentuálním podílem GC bází a vyšším procentuálním podílem AT bází. Minimální a maximální specifikace je nastavena tak, aby byla zajištěna analytická účinnost. |
| Průměrné mitochondriální pokrytí | ≥ 500 | Pokrytí mitochondriálního chromozomu. Minimální specifikace je nastavena tak, aby byl zajištěn limit detekce mitochondriálních SNV. |
| Procento bází Q30 | ≥ 85 | Míra kvality bází. Minimální specifikace je nastavena tak, aby byla zajištěna analytická účinnost. |
| Odhadovaná kontaminace vzorku | $\leq 0,005$ | Detekuje kontaminující čtení z jiných vzorků. Maximální specifikace je nastavena tak, aby byl zajištěn limit detekce mitochondriálních SNV (typ variant s nejvyšší citlivostí na kontaminaci). |
| Souhrnný výsledek kontroly kvality knihovny vzorků | PASS (Úspěšné) nebo FAIL (Neúspěšné) | V případě nesplnění požadavků kontroly kvality si prostudujte část Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132). |

Tabulka 5 Metriky kontroly kvality ploidy

| Metrika | Specifikace | Popis |
|-------------------------------------|--------------------------------|--|
| Zadaná ploidy pohlavních chromozomů | — | Pohlaví zadané operátorem během vytváření běhu (žena, muž, neznámé). |
| Odhad ploidy | — | Pohlavní ploidy odhaduje DRAGEN. |
| Souhrnný výsledek ploidy | CONCORDANT, DISCORDANT nebo ND | CONCORDANT (Shodné) označuje shodu mezi zadanou a odhadovanou pohlavní plodií. ND (Neurčeno) označuje pohlaví uvedené jako Neznámé nebo odhad jiný než XX nebo XY. U výsledků DISCORDANT (Neshodné) bylo během vytváření běhu zadáno nesprávné pohlaví nebo mohlo dojít k výměně vzorků. Viz část Řešení problémů v Příbalový leták k produktu TruSight pro analýzu celého genomu (dokument č. 200050132). |

Tabulka 6 Metriky pouze pro informaci

| Metrika | Popis |
|--|---|
| Medián délky insertů | Cíl je 450 bp, ale očekává se proměnlivost v závislosti na sekvenčním běhu a operátorovi. Přijatelné je rozmezí přibližně 360 až 550 bp. Trvalý provoz mimo tento rozsah může vést k vyššímu výskytu nevyhovujících vzorků. |
| Procento mapovaných čtení | Procento čtení, která se shodují s referenčním genomem. Může se snížit v reakci na kontaminaci nehumánní gDNA, špatnou kvalitu vzorku nebo příliš malou délku insertu, což vede k problémům s mapováním. |
| Procento čtení s doplňkovými zarovnáními | Procento čtení s mapováním, které se rozdělí na různá místa v referenčním genomu. |
| Procento duplicitních označených čtení | Očekávejte <20 %. Může být zvýšena, pokud je výtěžnost přípravy knihovny nízká nebo vkládání do fondu je menší než požadovaný objem, případně v reakci na problémy související se sekvenováním. |
| Procentuální podíl bází odstraněných metodou soft clipping – Čtení 1 | Užitečné při diagnostice hlavní příčiny selhání průměrného autozomálního pokrytí. |

| Metrika | Popis |
|--|---|
| Procentuální podíl bází odstraněných metodou soft clipping – Čtení 2 | Užitečné při diagnostice hlavní příčiny selhání průměrného autozomálního pokrytí. |
| Procentuální počet oříznutých bází – Čtení 1 | Užitečné při diagnostice hlavní příčiny selhání průměrného autozomálního pokrytí. |
| Procentuální počet oříznutých bází – Čtení 2 | Užitečné při diagnostice hlavní příčiny selhání průměrného autozomálního pokrytí. |

Informace o souboru přiřazení varianty

Soubory VCF

Soubory ve formátu přiřazení variant (*.vcf) obsahují informace o variantách nalezených na určitých pozicích v referenčním genomu a lze je nalézt v adresáři <Sample_ID>/Analysis.

Záhlaví souboru VCF obsahuje verzi formátu souboru VCF a verzi detektoru variant a uvádí poznámky použité ve zbytku souboru. Poslední řádek v záhlaví obsahuje nadpisy sloupců pro datové řádky. Každý z datových řádků souboru VCF obsahuje informace o jedné referenční pozici.

Všechny soubory VCF obsahují záhlaví s popisem výstupních sloupců a údaje o přiřazení variant ve sloupcích označených jako CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, SAMPLE. Definice hodnot sloupců se mohou u různých detektorů variant lišit.

Malé varianty a mSNV VCF

Výstup je uložen v souboru <Sample_ID>.annotated.hard-filtered.gvcf.gz v adresáři <Sample_ID>/Analysis.

Genomový soubor VCF (gVCF) obsahuje informace o variantách a pozicích určených jako homozygotní vůči referenčnímu genomu. U homozygotních oblastí soubor gVCF obsahuje statistické údaje, které ukazují, jak dobře čtení podporuje nepřítomnost variant nebo alternativních alel. Soubor gVCF obsahuje umělou <NON_REF> alelu. Čtení, která nepodporují referenci nebo jakékoli varianty, jsou přiřazeny alele <NON_REF>. DRAGEN používá tato čtení k určení, zda lze pozici označit jako homozygotní referenci,

místo toho, aby zůstala nepřirazená. Výsledné skóre představuje úroveň důvěryhodnosti homozygotního referenčního přiřazení podle stupnice Phred. V germinálním režimu je skóre FORMAT/GQ.

DRAGEN poskytuje filtrování variant po VCF na základě anotací přítomných v záznamech VCF. Tvrdé filtrování variant je popsáno níže. Vzhledem k povaze algoritmů sekundární analýzy DRAGEN, které zahrnují hypotézu o korelovaných chybách z jádra detektoru variant, má však plán lepší schopnosti při rozlišování skutečných variant od šumu, a proto se podstatně snižuje závislost na post-VCF filtrování.

TruSight Whole Genome Analysis Application poskytuje anotaci skóre spolehlivosti a úrovně spolehlivosti pro malé varianty, které lze použít k dalšímu zlepšení výkonnosti. Anotace úrovně spolehlivosti není filtrem kvality, a proto se přímo neodráží ve stavu kvality přiřazení variant. Proto lze vidět vyhovující přiřazení variant, která jsou nicméně anotována jako málo spolehlivá.

Tabulka 7 Záhlaví souboru VCF

| Záhlaví | Popis |
|---------|--|
| CHROM | Chromozom referenčního genomu. Chromozomy se zobrazují ve stejném pořadí jako referenční soubor FASTA. |
| POS | Jednobázová pozice varianty v referenčním chromozomu. Pro jednonukleotidové varianty (SNV) je tato pozice referenční bází s danou variantou. U indelů je tato pozice referenční bází bezprostředně předcházející dané variantě. |
| ID | Vždy . |
| REF | Název referenčního genotypu. Například delece jediného T je reprezentována jako referenční TT a alternativní T. Varianta jednoho nukleotidu A až T je reprezentována jako referenční A a alternativní T. |
| ALT | Alely, které se liší od referenčního čtení. Například inserce jediného T je reprezentována jako referenční A a alternativní AT. Varianta jednoho nukleotidu A až T je reprezentována jako referenční A a alternativní T. |
| QUAL | Skóre kvality podle stupnice Phred přiřazené detektorem variant. Vyšší skóre ukazuje na vyšší důvěryhodnost dané varianty a nižší pravděpodobnost chyb. U skóre kvality Q je odhadovaná pravděpodobnost chyby $10^{-(Q/10)}$. Například sada přiřazení Q30 má 0,1 % chybovost. Řada detektorů variant přiřazuje skóre kvality na základě svých statistických modelů, které jsou ve vztahu k pozorované chybovosti vysoké. |

Tabulka 8 Poznámky k souborům VCF

| Záhlaví | Popis |
|---------|---|
| FILTR | <ul style="list-style-type: none"> • PASS (Úspěšné) – Všechny filtry byly úspěšné. • DRAGENSnpHardQUAL – nastavte, pokud je hodnota true:QUAL <10,41 • DRAGENIndelHardQUAL – nastavte, pokud je hodnota true:QUAL <7,83 • LowDepth – nastavte, pokud je hodnota true:DP <= 1 • LowGQ – nastavte, pokud je hodnota true:GQ = 0 • PloidyConflict – přiřazení genotypu detektorem variant není konzistentní s chromozomovou ploidí • base_quality – místo je filtrováno, protože medián kvality bází alternativních čtení v tomto lokusu nespĺňuje prahovou hodnotu • filtered_reads – místo je filtrováno, protože byla odfiltrována příliš velká frakce čtení • fragment_length – místo je filtrováno, protože absolutní rozdíl mezi mediánem délky fragmentu alternativních čtení a mediánem délky fragmentu referenčních čtení v tomto lokusu překračuje prahovou hodnotu • low_af – frekvence alel nespĺňuje prahovou hodnotu • low_depth – místo je filtrováno, protože hloubka čtení je příliš nízká • low_frac_info_reads – místo je filtrováno, protože frakce informativních čtení je pod prahovou hodnotou • low_normal_depth – místo je filtrováno, protože hloubka čtení normálního vzorku je příliš nízká • long_indel – místo je filtrováno, protože délka indelu je příliš dlouhá • mapping_quality – místo je filtrováno, protože medián kvality mapování alternativních čtení v tomto lokusu nespĺňuje prahovou hodnotu • multiallelic – místo je filtrováno, protože více než dvě alely splňují LOD nádoru • non_homref_normal – místo je filtrováno, protože genotyp normálního vzorku není homozygotní referencí • no_reliable_supporting_read – místo je filtrováno, protože neexistuje spolehlivé podpůrné somatické čtení • panel_of_normals – zobrazeno alespoň v jednom vzorku ve vcf panelu normálů • read_position – místo je filtrováno, protože medián vzdáleností mezi začátkem/koncem čtení a tímto lokusem je pod prahovou hodnotou • RMxNRepeatRegion – místo je filtrováno, protože celá variantní alela nebo její část je opakováním reference • strand_artifact – místo je filtrováno z důvodu závažného vychýlení vlákna • str_contraction – místo je filtrováno kvůli podezření na chybu PCR, kde je alternativní alela o jednu opakovanou jednotku menší než reference • too_few_supporting_reads – místo je filtrováno, protože ve vzorku nádoru je příliš málo podpůrných čtení • weak_evidence – skóre somatických variant nespĺňuje prahovou hodnotu |

| Záhlaví | Popis |
|-----------|--|
| INFORMACE | <ul style="list-style-type: none"> • DB – členství v dbSNP • FS – p-hodnota podle stupnice Phred měřená pomocí Fisherova exaktního testu k detekci vychýlení vlákna • QD – důvěryhodnost/kvalita variant podle hloubky • R2_5P_bias – skóre založené na „mate bias“ a vzdálenosti od 5 prime konce. • SOR – symetrický poměr pravděpodobností 2x2 kontingenční tabulky pro detekci vychýlení vlákna • DP – přibližná hloubka čtení (informativní a neinformativní); některá čtení mohla být filtrována na základě mapq atd. • END – poloha zastavení intervalu • FractionInformativeReads – frakce informativních čtení z celkového počtu čtení • MQ – kvalita mapování RMS • MQRankSum – Z-skóre z Wilcoxonova sumárního testu kvality mapování čtení Alt oproti čtení Ref • ReadPosRankSum – Z-skóre z Wilcoxonova sumárního testu pořadí vychýlení pozice při čtení Alt oproti čtení Ref • SOMATIC – alespoň jedna varianta na této pozici je somatická • IENS – délky indelů pro každou variantu ALT. • SCORE – skóre spolehlivosti pro každý typ varianty přítomný v daném místě jako (typ varianty):(skóre spolehlivosti). • TIER – úroveň spolehlivosti pro každý typ varianty přítomný v daném místě jako (typ varianty):(úroveň spolehlivosti). |

| Záhlaví | Popis |
|---------|---|
| FORMAT | <p>Ve sloupci FORMAT (Formát) jsou uvedena pole oddělená dvojtečkami, například GT:GQ.</p> <ul style="list-style-type: none"> • AD – hloubky alel (počítají se pouze informativní čtení z celkových čtení) pro ref a alt alely v uvedeném pořadí • AF – alelové frakce pro alt alely v uvedeném pořadí • DP – přibližná hloubka čtení (čtení s MQ = 255 nebo se špatnými „mate“ jsou filtrována) • F1R2 – počet čtení v párové orientaci F1R2 podporující každou alelu • F2R1 – počet čtení v párové orientaci F2R1 podporující každou alelu • GP – posteriorní pravděpodobnosti podle stupnice Phred pro genotypy, jak jsou definovány ve specifikaci VCF • GQ – kvalita genotypu • GT – genotyp • ICNT – počet informativních čtení INDEL na základě referenčního modelu spolehlivosti • MB – statistiky komponent vztažené na vzorek pro detekci „mate bias“ • MIN_DP – minimální DP pozorovaný v bloku GVCF • PL – normalizované pravděpodobnosti podle stupnice Phred pro genotypy, jak jsou definovány ve specifikaci VCF • PRI – předchozí pravděpodobnosti podle stupnice Phred pro genotypy • PS – informace o ID fyzického fázování, kde každé jedinečné ID v daném vzorku (ale nikoli napříč vzorky) spojuje záznamy v rámci skupiny fázování • SB – statistiky komponent vztažené na vzorek, které tvoří Fisherův exaktní test pro detekci vychýlení vlákna • SPL – normalizované pravděpodobnosti SNP podle stupnice Phred na základě referenčního modelu spolehlivosti • SQ – somatická kvalita |
| VZOREK | Sloupec vzorku uvádí hodnoty specifikované ve sloupci FORMÁT. |

VCF pro varianty počtu kopií

Fáze cílových počtů je první fází zpracování plánu CNV systémem DRAGEN, která vytváří soubor `<Sample_ID>.target.counts.gz`, poté se provede GC Bias Correction (Korekce zkreslení GC) a vytvoří se soubor `*.target.counts.gc-corrected.gz`. Fáze normalizace vytváří soubor `*.tn.tsv.gz`. Hostitelský software systému DRAGEN vytváří mnoho mezisouborů. `*.seg.called.merged` je konečný soubor přiřazení, který obsahuje amplifikační a deleční události. Spolu se segmentovým souborem DRAGEN odesílá přiřazení ve standardním formátu VCF. Výstup je uložen v souboru `<Sample_ID>.cnv.vcf.gz` v adresáři `<Sample_ID>/Analysis`.

Definice sloupců specifických pro detektor CNV:

Sloupec POS je počáteční poloha varianty. Podle specifikace VCF, pokud je některá z alel ALT symbolickou alelou, například , pak se vyžaduje výplňová báze a POS označuje souřadnici báze předcházející polymorfismu. Všechny souřadnice ve VCF jsou založeny na 1.

Sloupec ID slouží k reprezentaci události. Pole ID kóduje typ události a souřadnice události.

Sloupec REF obsahuje N pro všechny události CNV.

Sloupec ALT specifikuje typ události CNV. Jelikož VCF obsahuje pouze události CNV, používá se pouze položka DEL nebo DUP.

Sloupec QUAL obsahuje odhadované skóre kvality pro událost CNV, které se používá při tvrdém filtrování.

Sloupec FILTER obsahuje hodnotu PASS (Úspěšné), pokud událost CNV projde všemi filtry, jinak sloupec obsahuje název neúspěšného filtru.

Sloupec INFO obsahuje informace o události. Položka REFLLEN označuje délku události. Položka SVTYPE je vždy CNV. Položka END označuje koncovou pozici události.

Pole FORMAT (Formát) jsou popsána v záhlaví.

- GT – genotyp
- SM – poměr lineární kopie střední hodnoty segmentu
- CN – odhadované číslo kopie
- BC – počet dílů v regionu
- PE – počet nesprávně spárovaných koncových čtení na počátečních a koncových bodech zlomu

VCF pro opakování

ExpansionHunter provede nové zarovnání čtení na základě grafu sekvencí, která pochází z každého opakování cíle a jeho okolí. ExpansionHunter pak genotypuje délku repetice v každé alele na základě těchto zarovnání grafů.

Další informace a analýzy jsou k dispozici v následujících pracích týkajících se systému ExpansionHunter:

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al., *ExpansionHunter: A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

Katalog variant TruSight Whole Genome Analysis Application STR obsahuje specifikace repetitivních způsobujících onemocnění, které se nacházejí v genech AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B a TBP.

Výsledky opakované genotypizace jsou vyvedeny jako samostatný soubor VCF, který poskytuje délku každé alely u každé přiřaditelné repetice definované v katalogovém souboru specifikace repetice. Název tohoto souboru je <Sample_ID>.repeats.gz a lze jej najít v adresáři <Sample_ID>/Analysis.

Některé sloupce jsou specifické pro detektor expanze repetice:

Tabulka 9 Základní pole VCF

| Pole | Popis |
|-------|--|
| CHROM | Identifikátor chromozomu |
| POS | Pozice první báze před oblastí repetice v referenci |
| ID | Vždy . |
| REF | Referenční báze na pozici POS |
| ALT | Seznam opakovaných alel ve formátu <STRn>. N je počet opakovaných jednotek. |
| QUAL | Vždy . |
| FILTR | Filtr LowDepth se použije, když je celková hloubka lokusu pod 10násobkem nebo počet čtení, která přesahují jeden nebo oba zlomy, je pod 5. |

Tabulka 10 Další pole s informacemi

| Pole | Popis |
|-------|--|
| END | Poloha poslední báze před oblastí repetice v referenci |
| REF | Číslo referenční kopie |
| REPID | ID varianty z katalogu variant |
| RL | Délka reference v bp |
| RU | Jednotka opakování v referenční orientaci |
| VARID | ID varianty z katalogu variant |

Tabulka 11 Pole GENOTYPE (pro každý vzorek)

| Pole | Popis |
|------|--|
| AD | Hloubky alel pro referenční a alternativní alely v uvedeném pořadí |
| ADFL | Počet doprovodných čtení konzistentních s alelou |
| ADIR | Počet čtení v rámci repetice konzistentních s alelou |
| ADSP | Počet překlenovacích čtení konzistentních s alelou |
| DST | Výsledky (+ detekováno, - nedetekované, ? neurčeno) testu reprezentovaného variantou |
| GT | Genotyp |

| Pole | Popis |
|-------|--|
| LC | Pokrytí lokusu |
| REPCI | Interval spolehlivosti pro REPCN |
| REPCN | Počet opakovaných jednotek překlenutých alelou |
| RPL | Poměr logaritmické pravděpodobnosti pro přítomnost referenční alely |
| SO | Typ čtení, která podporují alelu. Hodnoty mohou být SPANNING, FLANKING nebo INREPEAT. Tyto hodnoty udávají, zda čtení přesahují repetici, jsou s ní souběžná nebo jsou plně obsažena v repetici. |

Soubor `<Sample_ID>.repeats.vcf.gz` obsahuje výstup SMN spolu s veškerými cílenými repeticemi. Výstup SMN je reprezentován jako jediné přiřazení SNV na pozici ovlivňující splice v SMN1 (NM_000344.3:c.840C/T) se stavem spinální svalové atrofie (SMA) v následujících vlastních polích.

Tabulka 12 Výsledky SMA ve výstupním souboru repeats.vcf

| Pole | Popis |
|-------|---|
| VARID | SMN označuje přiřazení SMN. |
| GT | Přiřazení genotypu na této pozici pomocí modelu normálního (diploidního) genotypu. |
| DST | Přiřazení stavu SMA: + označuje detekované - označuje nedetekované ? označuje neurčené |
| AD | Celkové počty čtení, které podporují alelu C a T. |
| RPL | Poměr pravděpodobnosti Log10 mezi neovlivněnými a ovlivněnými modely. Kladné skóre naznačuje, že neovlivněný model je pravděpodobnější. |

ROH BED

Oblasti homozygoty (ROH) jsou detekovány v rámci detektorů malých variant. Detekční program detekuje a zobrazuje běhy homozygoty z celogenomových přiřazení na autozomálních lidských chromozomech. Pohlavní chromozomy jsou ignorovány, pokud pohlavní karyotyp vzorku není XX, jak je určeno pomocí nástroje pro odhad ploidie. Výstup ROH umožňuje následným nástrojům vyhledávat a předpovídat příbuzenské vztahy mezi rodiči probanda.

Oblast je definována jako po sobě jdoucí přiřazení variant na chromozomu bez velké mezery mezi těmito variantami. Jinými slovy, oblasti jsou rozbity chromozomem nebo velkými mezerami bez přiřazení SNV. Velikost mezery je nastavena na 3 Mbáze.

Detektor ROH vytvoří výstupní soubor ROH s názvem `<Sample_ID>.roh.bed` v adresáři `<Sample_ID>/Analysis`. Každý řádek představuje jednu oblast homozygoty. Soubor `.bed` obsahuje následující sloupce:

| Chromozom | Začátek | Konec | Skóre | #Počet homozygotních | #Počet heterozygotních |
|-----------|---------|-------|-------|----------------------|------------------------|
|-----------|---------|-------|-------|----------------------|------------------------|

Kde

- Skóre je funkcí počtu homozygotních a heterozygotních variant, kde každá homozygotní varianta zvyšuje skóre o 0,025 a každá heterozygotní varianta snižuje skóre o 0,975.
- Počáteční a koncová pozice jsou polootevřený interval založený na 0.
- #Homozygous (Počet homozygotních) je počet homozygotních variant v regionu.
- #Heterozygous (Počet heterozygotních) je počet heterozygotních variant v regionu.

Detektor také vytvoří soubor metrik s názvem `<Sample_ID>.roh_metrics.csv`, který uvádí počet velkých ROH a procento SNP ve velkých ROH (>3 MB).

Metriky pro odhad ploidie

Ve výchozím nastavení běží nástroj pro odhad ploidie. Nástroj pro odhad ploidie používá čtení z mapovače/zarovnávače k výpočtu hloubky sekvenování pokrytí pro každý autozom a alozom v lidském genomu. Pohlavní karyotyp vzorku je pak odhadnut pomocí poměrů mediánu pokrytí pohlavního chromozomu k mediánu autozomálního pokrytí. XX nebo XY a CONCORDANT, DISCORDANT nebo ND (neurčeno) ve srovnání s poskytnutými údaji o vzorcích jsou uvedeny v konsolidované zprávě. Podrobné výsledky, včetně každého normalizovaného mediánu pokrytí na contig, jsou uvedeny v souboru `<Sample_ID>.ploidy_estimation_metrics.csv`.

Soubory FASTQ

FASTQ (*.fastq.gz, *.fastq.ora) je textový formát souboru obsahující přiřazení bází a hodnoty kvality na jednotlivé čtení. Každý soubor obsahuje následující informace:

- Identifikátor vzorku
- Sekvence
- Znaménko plus (+)
- Skóre kvality Phred v kódovaném formátu ASCII + 33

Software vygeneruje jeden soubor FASTQ pro každý vzorek, čtení a řadu. Například pro každý vzorek v cyklu s párovým koncem software vygeneruje dva soubory FASTQ: jeden pro čtení 1 a jeden pro čtení 2. Kromě těchto souborů FASTQ pro vzorky vygeneruje software také dva soubory FASTQ pro každou řadu obsahující všechny neznámé vzorky. Soubory FASTQ pro Index Read 1 (Indexové čtení 1) a Index Read 2 (Indexové čtení 2) se negenerují, protože sekvence je zahrnuta v záhlaví každého záznamu FASTQ. Formát názvu souboru je sestaven z polí zadaných v seznamu vzorků a používá formát pojmenování souboru `<Sample_ID>_S#_L00#_R#_001.fastq.gz`

Soubory FASTQ jsou uloženy v adresáři `<Sample_ID>/Conversion`. V adresáři FASTQ složky analýzy se nachází adresář Logs (Protokoly) s protokoly převodu BCL do FASTQ a adresář Reports (Zprávy),

který obsahuje různé soubory s metrikami čtení a soubor `SampleSheet.csv` použitý pro převod do FASTQ. Soubory FASTQ z neurčených čtení se nacházejí v adresáři `Undetermined/Conversion` (Neurčené/Konverze) ve složce Analysis (Analýza).

Identifikátor vzorku je naformátován následovně:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber
```

Příklad:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
```

```
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
```

```
+
```

```
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

Soubory CRAM

Soubory Compressed Reference-oriented Alignment Map (komprimovaná referenčně orientovaná mapa zarovnání) nebo-li CRAM (*.cram) jsou uloženy v adresáři `<Sample_ID>/Analysis` a obsahují záhlaví a záznamy o zarovnání vzhledem ke genomickému referenčnímu souboru používanému během zarovnání. Cesta k referenčnímu souboru je uvedena v souboru `<Sample_ID>/Analysis/<Sample_ID>-replay.json` jako parametr `--ht-reference`, nastavený na `hg38.fa`.

Soubory CRAM obsahují sekci záhlaví a sekci zarovnání:

- **Header** (Záhlaví) – obsahuje informace o celém souboru, jako je název vzorku, délka vzorku a metoda zarovnání. Zarovnání (Alignments) jsou v sekci zarovnání spojena s konkrétními informacemi v sekci záhlaví.
- **Alignments** (Zarovnání) – obsahuje název čtení, sekvenci čtení, kvalitu čtení, informace o zarovnání a vlastní tagy. Název čtení zahrnuje chromozom, počáteční souřadnici, kvalitu zarovnání a řetězec deskriptoru shody.

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

- AS: Kvalita zarovnání párového konce.
- RG: Skupina čtení, která označuje počet čtení pro konkrétní vzorek.
- BC: Tag s čárovým kódem, který označuje ID demultiplexovaného vzorku spojené se čtením.
- SM: Kvalita zarovnání jednoho konce.
- XC: Shoda řetězce deskriptoru.
- XN: Značka názvu amplikonu, která zaznamenává ID amplikonu spojeného se čtením

Chcete-li zobrazit záznamy o zarovnání, lze `samtools` použít jako `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram`.

Generuje se také indexový soubor a soubor s kontrolním součtem.

Technická pomoc

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

Web: www.illumina.com

E-mail: techsupport@illumina.com

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

Dokumentace k produktu – k dispozici ke stažení z webu support.illumina.com.



Illumina, Inc.
5200 Illumina Way
San Diego, Kalifornie 92122, Spojené státy
americké
+1 800 809 ILMN (4566)
+1 858 202 4566 (mimo Severní Ameriku)
techsupport@illumina.com
www.illumina.com



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

Australský zadavatel

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

URČENO K DIAGNOSTICE IN VITRO.

© 2024 Illumina, Inc. Všechna práva vyhrazena.

illumina[®]