

DRAGEN for Illumina DNA Prep with Enrichment Dx

Produktová dokumentácia prístroja NovaSeq 6000Dx

História revízií

Dokument	Dátum	Popis zmeny
200014776 v02	September 2022	Opravený formát súboru manifest z textu (*.txt) na BED (*.bed) v pokynoch na vytvorenie chodu. Opravené konsenzuálne súbory VCF na súbory VCF v sekcii výstupu analýzy.
200014776 v01	August 2022	Pridané: Časť Nastavenia. Časť Systematické filtrovanie hluku. Pokyny na vytvorenie chodu, aby obsahoval podrobnejšie informácie boli aktualizované. Boli opravené preklepy a gramatické chyby. Bolo špecifikované, že pokyny sú určené pre aplikáciu pri použití s prístrojom NovaSeq 6000Dx. Informácie o obsahu výstupného súboru VCF boli aktualizované.
200014776 v00	Marec 2022	Úvodné vydanie.

Tento dokument a jeho obsah sú vlastníctvom spoločnosti Illumina, Inc. a jej pridružených spoločností (ďalej len „Illumina“) a sú určené výlučne na zmluvné použitie u zákazníka v súvislosti s používaním produktu (produktov) opísaného (opísaných) v tomto dokumente a na žiadny iný účel. Tento dokument a jeho obsah sa nesmú používať ani šíriť na žiadny iný účel a/alebo inak poskytovať, zverejňovať alebo reprodukovať akýmkoľvek spôsobom bez predchádzajúceho písomného súhlasu spoločnosti Illumina. Spoločnosť Illumina týmto dokumentom neposkytuje žiadnu licenciu na základe patentu, ochrannej známky, autorských práv alebo práv podľa zvykového práva, či podobných práv tretích strán.

Pokyny v tomto dokumente musia byť prísne a výslovne dodržiavané kvalifikovaným a riadne vyškoleným personálom, aby sa zabezpečilo správne a bezpečné používanie tu popísaného výrobku (výrobov). Pred použitím takéhoto výrobku (výrobov) je nutné prečítať si celý obsah tohto dokumentu s porozumením.

NEPREČÍTANIE VŠETKÝCH POKYNOV TU OBSIAHNUÝCH A ICH VÝSLOVNÉ NEDODRŽANIE MÔŽE MAŤ ZA NÁSLEDOK POŠKODENIE VÝROBKU (VÝROBKOV), ZRANENIE OSOBY VRÁTANE POUŽÍVATEĽOV ALEBO INÝCH OSÔB, POŠKODENIE ĎALŠIEHO MAJETKU A ZRUŠENIE PLATNOSTI ZÁRUKY VZŤAHUJÚCEJ SA NA VÝROBK (VÝROBKY).

SPOLOČNOSŤ ILLUMINA NEPREBERÁ ŽIADNU ZODPOVEDNOSŤ VYPLÝVAJÚCU Z NEBEZPEČNÉHO POUŽITIA TU POPÍSANÉHO VÝROBKU (VÝROBKOV) (VRÁTANE SÚČASTÍ ALEBO SOFTVÉRU).

© 2022 Illumina, Inc. Všetky práva vyhradené.

Všetky ochranné známky sú vlastníctvom spoločnosti Illumina, Inc. alebo príslušných vlastníkov. Informácie o konkrétnych ochranných známkach nájdete na stránke www.illumina.com/company/legal.html.

Obsah

História revízií	ii
Základné informácie	1
Metódy analýzy	1
Vytvorenie chodu	4
Nastavenia	5
Výstupy analýzy	7
Súbory FASTQ	8
Súbory BAM	8
Súbory VCF	9
Zobrazenie výsledkov analýzy	14
Technická pomoc	15

Základné informácie

DRAGEN™ for Illumina® DNA Prep with Enrichment Dx vykonáva demultiplexovanie, generovanie súborov FASTQ, mapovanie čítania a zarovnanie s referenčným genómom a analýzou variantov v závislosti od vybraného pracovného postupu analýzy.

Metódy analýzy

DRAGEN for Illumina DNA Prep with Enrichment Dx vykonáva demultiplexovanie, generovanie súborov FASTQ, mapovanie čítania a zarovnanie s referenčným genómom v závislosti od vybraného pracovného postupu:

- Generovanie súborov FASTQ
- Generovanie súborov Germline FASTQ a VCF
- Generovanie somatických súborov FASTQ a VCF

Generovanie súborov FASTQ

Zostavené sekvencie sú zapísané do súborov FASTQ na každú vzorku. Súbor FASTQ sú textové súbory, ktoré obsahujú údaje a skóre kvality sekvenovania iba pre jednu vzorku. Pre každú vzorku sa vygenerujú samostatné súbory FASTQ na každý pruh prietokového článku a na každé čítanie sekvenovania. Názov vzorky, ktorý je uvedený počas nastavenia chodu, je uvedený v názve súboru FASTQ. Súbor FASTQ predstavujú primárny vstup na zarovnanie. Prvým krokom pri generovaní súborov FASTQ je demultiplexovanie. Demultiplexovanie priraduje klastre, ktoré prejdú filtrom do vzorky, porovnaním každej sekvencie čítania indexu so sekvenciami indexov špecifikovanými v rámci chodu. V rámci tohto kroku sa nezohľadňujú žiadne kvalitatívne hodnoty. Identifikácia načítania indexov prebieha podľa nasledujúcich krokov:

- Vzorky sú očíslované od 1 na základe poradia, v akom sú uvedené v rámci chodu.
- Číslo vzorky 0 je vyhradené pre klastre, ktoré neboli priradené k vzorke.
- Klastre sa priradujú k vzorke vtedy, keď sa sekvencia indexu presne zhoduje, alebo vtedy, ak načítanie indexu vykazuje max. jednu nezgodu.

Softvér obsahuje komprimáciu ORA na komprimáciu súborov FASTQ. Pri použití formátu ORA (*.ora) sa kontrolný súčet md5 obsahu súborov FASTQ zachová po cykle komprimácie a dekomprimácie za účelom predchádzania strát.

Mapovanie & zarovnanie DNA

Prvou fázou mapovania je generovanie zdrojov z čítania a potom hľadanie presných zhôd v referenčnom genóme. Tieto výsledky sú potom spresnené spustením úplných Smith-Watermanových zarovnaní na miestach s najvyššou hustotou zhôd zdrojov. Tento dobre zdokumentovaný algoritmus

funguje tak, že porovnáva každú pozíciu čítania so všetkými kandidátskymi pozíciami referenčných hodnôt. Tieto porovnania zodpovedajú matici potenciálnych zoradení medzi čítanou a referenčnou hodnotou. Pre každú z týchto kandidátskych pozícií Smith-Watermanove zarovnanie vygeneruje skóre, ktoré sa používa na vyhodnotenie toho, či sa najlepšie zarovnanie prechádzajúce touto maticovou bunkou dosiahne zhodou alebo nezhodou nukleotidov (diagonálny pohyb), vymazaním (horizontálny pohyb) alebo vložením (vertikálny pohyb). Zhoda medzi čítanou a referenčnou hodnotou poskytuje bonus za skóre a nezhoda alebo indel ukladá trest. Vybraná je celková cesta cez maticu s najvyšším skóre.

Špecifické hodnoty zvolené pre skóre v tomto algoritme naznačujú ako v rámci zarovnania s viacerými možnými interpretáciami vyvážiť možnosť indelu na rozdiel od jedného alebo viacerých SNP alebo predvoľbu zarovnania bez orezania. Predvolené hodnoty hodnotenia DRAGEN sú primerané na zarovnanie čítania strednej dĺžky s celým ľudským referenčným genómom pre aplikácie analýzy variantov. Akýkoľvek súbor parametrov Smith-Watermanovho hodnotenia predstavuje nepresný model genómových mutácií a chýb sekvenovania. Pre niektoré aplikácie môžu byť vhodnejšie rôzne ladené skórovacie hodnoty zarovnania.

Analýza variantu DRAGEN Germline

DRAGEN Germline Small Variant Caller vykonáva mapované a zarovnané čítania DNA ako vstup a analyzuje SNP a indel prostredníctvom kombinácie stĺpcovej detekcie a lokálnej zostavy haplotypov *de novo*.

Analyzovateľné referenčné oblasti sú najprv identifikované s dostatočným pokrytím zarovnania. V rámci týchto referenčných oblastí rýchlym skenom zoradených čítaní sa identifikujú aktívne oblasti, ktoré sú vycentrované okolo nahromadených stĺpcov s dôkazom variantu. Aktívne oblasti sú vyplnené dostatkom kontextu na pokrytie významného nereferenčného obsahu v blízkosti. Ak existujú dôkazy o indeloch, aktívne oblasti dostanú dodatočnú výplň.

Zarovnané čítania sú orezané v každej aktívnej oblasti a zostavené do grafu De Bruijn. Okraje orezaných čítaní sú vážené počtom pozorovaní, pričom referenčná sekvencia bude opornou hodnotou. Po určitom vyčistení a zjednodušení grafu sa všetky cesty typu source-to-sink extrahujú ako kandidátske haplotypy. Každý haplotyp je Smith-Watermanovým algoritmom zarovnaný s referenčným genómom, aby sa identifikovali varianty, ktoré predstavuje. Tento súbor udalostí môže byť rozšírený detekciou na základe pozície. Pre každý pár haplotypu čítania sa pravdepodobnosť $P(r|H)$ pozorovania čítania, za predpokladu, že haplotyp je skutočnou východiskovou vzorkou, odhaduje pomocou párového skrytého Markovho modelu (HMM).

Skenovanie podľa referenčnej polohy nad aktívnou oblasťou, kandidátske genotypy sú vytvorené z diploidných kombinácií variantných udalostí (SNP alebo indel). Pre každú udalosť (vrátane referencie) sa podmienená pravdepodobnosť $P(r|e)$ pozorovania každého prekrývajúceho sa čítania odhadne ako maximálna $P(r|H)$ pre haplotypy podporujúce udalosť. Tieto sa skombinujú do podmienenej pravdepodobnosti $P(r|e1e2)$ pre genotyp (pár udalostí) a vynásobia sa, aby sa získala podmienená pravdepodobnosť $P(R|e1e2)$ pozorovania celého čítania. Pomocou Bayesovho vzorca sa vypočíta posteriórna pravdepodobnosť $P(e1e2|R)$ každého diploidného genotypu a vykoná sa analýza víťaza.

V režime gVCF, ktorý sa používa na škálovateľné predkladanie viacvzorkového variantu, možno spustiť DRAGEN Germline Small Variant Caller pre každú vzorku, aby sa vygeneroval pomocný súbor analýzy genómových variantov (gVCF). Režim gVCF sa potom môže použiť na efektívne spoločné genotypovanie viacerých vzoriek, čo umožňuje rýchle postupné spracovanie vzoriek a škálovanie na veľké kohorty.

Pretože DRAGEN Germline Small Variant Caller má algoritmy, ktoré umožňujú efektívne rozlíšiť korelované chyby od skutočných variantov, pravidlá filtrovania sú veľmi jednoduché.

Analýza somatického variantu DRAGEN

DRAGEN Somatic Small Variant Caller berie zmapované a zarovnané čítania DNA ako vstup a analyzuje SNV a indely prostredníctvom lokálnej *de novo* zostavy haplotypov v aktívnej oblasti.

Analyzovateľné referenčné oblasti sú najprv identifikované s dostatočným pokrytím zarovnania. V rámci týchto referenčných oblastí skenovanie triedených hodnôt identifikuje aktívne oblasti, ktoré sú sústredené okolo nahromadených stĺpcov s dôkazom variantu v údajoch o nádore. Aktívne oblasti sú vyplnené dostatkom kontextu na pokrytie významného nereferenčného obsahu v blízkosti. Ak existujú dôkazy o indeloch, aktívne oblasti dostanú dodatočnú výplň.

Zarovnané čítania sú orezané v každej aktívnej oblasti a zostavené do grafu De Bruijn. Okraje orezaných čítaní sú vážené počtom pozorovaní, pričom referenčná sekvencia bude opornou hodnotou. Po určitom vyčistení a zjednodušení grafu sa všetky cesty typu source-to-sink extrahujú ako kandidátske haplotypy. Každý haplotyp je Smith-Watermanovým algoritmom zarovnaný s referenčným genómom, aby sa identifikovali varianty, ktoré predstavuje. Pre každý pár haplotyp čítania sa odhaduje pravdepodobnosť $P(r|H)$ pozorovania čítania pomocou párového skrytého Markovovho modelu (HMM) za predpokladu, že haplotyp je skutočnou východiskovou vzorkou.

Na určenie skóre TLOD DRAGEN Somatic Small Variant Caller najprv skenuje podľa referenčnej pozície každú kandidátsku somatickú udalosť, ako aj referenčnú udalosť v aktívnej oblasti. Podmienená pravdepodobnosť $P(r|e)$ pozorovania každého prekryvajúceho sa čítania sa odhaduje ako maximálne $P(r|H)$ pre haplotypy podporujúce udalosť. Tieto sa spájajú do podmienenej pravdepodobnosti $P(r|E)$ pre hypotézu udalosti, E, zahŕňajúcej zmes referenčnej a kandidátskej somatickej alely v rozsahu možných frekvencií alel a sú vynásobené, aby sa získala podmienená pravdepodobnosť $P(R|E)$ pozorovania všetkých nazhromaždených čítaní. Odtiaľ sa skóre TLOD vypočíta ako dôkaz, že alela ALT je prítomná vo vzorke nádoru na danom mieste.

Vytvorenie chodu

Pomocou nasledujúcich krokov nastavte chod v Správca chodu spoločnosti Illumina NovaSeq 6000Dx alebo pomocou prehliadača v sieťovom počítači. Údaje o vzorke je možné zadať manuálne alebo importovaním hárku vzoriek.

Nastavenia aplikácie & chodu

1. Na obrazovke Runs (Chody) vyberte **Create Run** (Vytvoriť chod).
2. Vyberte aplikáciu DRAGEN for Illumina DNA Prep with Enrichment Dx a potom možnosť **Next** (Ďalej).
3. Na obrazovke Run Settings (Nastavenia chodu) zadajte názov chodu. Zadajte názov chodu, ktorým sa bude označovať chod od sekvenovania po analýzu.
4. **[Voliteľné]** Na zjednodušenie označenia chodu zadajte opis chodu.
5. Uistite sa, že zvolená prípravná súprava knižnice je prípravná súprava knižnice Illumina DNA Prep with Enrichment Dx.
6. Vyberte požadovanú súpravu adaptéra indexu.
7. Zadajte dĺžku čítania.
Čítanie 1 a čítanie 2 majú predvolenú hodnotu 151 cyklov.
Index 1 a index 2 majú fixnú hodnotu 10 cyklov.
8. **[Voliteľné]** Zadajte ID skúmavky knižnice.
9. Vyberte možnosť **Next** (Ďalej).

Údaje o vzorke

Do tabuľky na obrazovke Sample Data (Údaje o vzorke) zadajte informácie o vzorke manuálne. Prípadne vyberte možnosť **Import Samples** (Importovať vzorky) na nahranie informácií o vzorke. Informácie o importovaní informácií o vzorke nájdete v časti [Importovať vzorky na strane 5](#).

Manuálne zadanie vzoriek

1. Do poľa Sample ID (ID vzorky) zadajte jedinečné ID vzorky.
2. Vyberte možnosť **Plate - Well Position** (Doštička – poloha jamky) a vyberte polohu jamky.
Polia indexu i7, indexu 1, indexu i5 a indexu 2 sa vyplnia automaticky.
3. **[Voliteľné]** Zadajte názov knižnice.
4. Pridajte riadky a podľa potreby opakujte kroky 1–3, kým do tabuľky nepridáte postupne všetky vzorky.
5. Vyberte možnosť **Next** (Ďalej).

Importovať vzorky

Šablóna (*.csv) je dostupná na stiahnutie na obrazovke Sample Data (Údaje o vzorke) pri plánovaní chodu Správca chodu spoločnosti Illumina pomocou prehliadača na sieťovom počítači.

1. Pre stiahnutie prázdneho súboru CSV kliknite na možnosť **Download Template** (Stiahnuť šablónu).
2. Zo súboru CSV zadajte informácie o vzorke a súbor uložte.
Súbor CVS hárku vzoriek obsahuje nasledujúce stĺpce údajov: Sample ID (ID vzorky), Plate - Well Position (Doštička – poloha jamky), názov **Optional** (Voliteľné) knižnice.
3. Vyberte možnosť **Import Samples** (Importovať vzorky) a prejdite na umiestnenie súboru CSV.
4. Vyberte možnosť **Next** (Ďalej).

Nastavenia analýzy

1. Vyberte požadovaný pracovný postup analýzy:
 - Generovanie súborov FASTQ
 - Generovanie súborov Germline FASTQ a VCF pre pracovný postup germline
 - Generovanie súborov Somatic FASTQ a VCF pre pracovný postup somatic
2. **[Voliteľné]** V prípade potreby vyberte možnosť **Generate ORA compressed FASTQs** (Generovať ORA komprimované súbory FASTQ) na povolenie komprimácie FASTQ ORA.
3. **[Pracovné postupy generovania súborov VCF]** Vyberte možnosť **Manifest File Selection** (Výber súboru manifest) z rozbaľovacej ponuky.
Súbor manifest je potrebný vstup pre DRAGEN for Illumina DNA Prep with Enrichment Dx. Manifest je súbor BED (*.bed), ktorý obsahuje názvy a umiestnenia cieľových referenčných oblastí.
4. **[Pracovné postupy generovania súborov Somatic FASTQ a VCF]** Vyberte možnosť **Noise File Selection** (Výber súboru šumu) z rozbaľovacej ponuky.
Na filtrovanie systematického šumu je možné špecifikovať súbor BED s hladinou šumu špecifickou pre dané pracovisko. Ďalšie informácie nájdete v časti [Filtrovanie šumu na strane 6](#).
5. Vyberte možnosť **Next** (Ďalej).

Kontrola chodu

1. Na obrazovke Review (Kontrola) skontrolujte informácie zadané na obrazovkách Run Settings (Nastavenia cyklu), Sample Data (Údaje o vzorke) a Analysis Settings (Nastavenia analýzy).
2. Vyberte možnosť **Save** (Uložiť).
Cyklus sa uloží na kartu Planned (Plánované) na obrazovke Runs (Chody).

Nastavenia

Ak chcete zobrazit' aktuálne nastavenia a zmenit' ich, vyberte aplikáciu na obrazovke Applications (Aplikácie).

Konfigurácia

Na konfiguračnej obrazovke sa zobrazujú nasledujúce nastavenia aplikácie:

- **Library Prep Kits** (Súpravy na prípravu knižnice) – zobrazuje predvolenú súpravu na prípravu knižnice v rámci aplikácie. Toto nastavenie nie je možné zmeniť.
- **Index Adapter Kits** (Súpravy indexových adaptérov) – zobrazuje predvolenú súpravu indexových adaptérov v rámci aplikácie. Toto nastavenie nie je možné zmeniť.
- **Read lengths** (Dĺžka čítania)– dĺžka čítania je pre aplikáciu predvolene nastavená na 151, ale môže sa zmeniť počas vytvárania chodu.
- **Manifest and Noise Files** (Súbory manifest a šumu) – nahrajte a zmeňte nastavenia súborov manifest a šumu.
 - Ak chcete nahrať súbory na použitie v analýze, vyberte možnosť **Upload File** (Nahrať súbor).
 - Stlačením **Default** (predvoleného) tlačidla nastavíte súbor ako predvolený súbor manifest alebo hluku vybraný počas vytvárania chodu, keď je vybraná aplikácia.
 - Začiarknutím políčka **Enabled** (Aktivované) nastavte súbor, ktorý sa má zobraziť v rozbaľovacej ponuke počas vytvárania chodu.

Povolenia

Začiarkavacie políčka na obrazovke Permissions (Povolenia) použite na správu prístupu používateľa v rámci aplikácie.

Filtrovanie šumu

Pri použití somatického pracovného postupu je k dispozícii systematické filtrovanie šumu. Filter je možné použiť v režime Tumor-Normal (Nádor-bežný), ale je obzvlášť užitočný pri chodoch Tumor-Only (Len nádor), pri ktorých nie je dostupná zhoda pri normal (bežný).

Systematický hluk BED by sa mal generovať z bežných vzoriek. Odporúča sa vytvárať systematické súbory hluku, ktoré sú špecifické pre prípravu knižnice, systém sekvenovania a panel. Na vytvorenie súboru hluku sa odporúča použiť približne 50 bežných vzoriek.

Výstupy analýzy

DRAGEN for Illumina DNA Prep with Enrichment Dx uloží nasledujúce informácie do priečinka analýzy. Iba pracovné postupy germline a somatic generujú PDF súbor.

- Použitý súbor manifestu
- Verzia softvéru
- ID vzorky
- Celkový počet zarovnaných čítaní
- Percentuálny podiel zarovnaných čítaní na každú vzorku
- Počet SNV stanovených na každú vzorku
- Počet indelov stanovených na každú vzorku
- Štatistika pokrytia

Súbory s výstupmi analýzy

Aplikácia vygeneruje nasledujúce výstupné súbory. Presné vygenerované súbory závisia od toho, ktorý pracovný postup analýzy bude použitý. Výstupné súbory sa nachádzajú v priečinku analýzy.

Výstupný súbor	Popis
Súbory FASTQ (*.fastq.gz alebo *.fastq.ora)	Medzisúbory obsahujúce primárne analýzy báz so skóre kvality. Súbory FASTQ predstavujú primárny vstup na krok zarovnaní. Ak je vybraná komprimácia ORA, v názve súboru to bude uvedené.
Zarovnanie súborov BAM (*.bam)	Obsahujú zarovnané čítania danej vzorky.
Genómové VCF súbory (*.gvcf.gz)	Obsahujú genotyp pre každú pozíciu, či už stanovenú ako variant alebo ako referencia.
Súbory VCF (*.vcf.gz)	Obsahujú varianty stanovené v každej pozícii.
Správa o metrike chodu (*.csv)	Obsahuje metriky kvality chodu vrátane celkovej výnosnosti a skóre Q30.

Súbory FASTQ

Súbor FASTQ (*.fastq.gz, *.fastq.ora) je textový formát súboru obsahujúci primárne analýzy báz a hodnoty kvality každého čítania. V každom súbore sú uvedené nasledujúce informácie:

- Identifikátor vzorky
- Sekvencia
- Znak plus (+)
- Phred skóre kvality v kódovanom formáte ASCII + 33

Identifikátor vzorky je naformátovaný nasledovne.

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

Súbory BAM

Súbor BAM (*.bam) je komprimovaná binárna verzia súboru SAM (mapa zarovnaní sekvencie), ktorá sa používa na zobrazenie zarovnaných sekvencií do 128 Mb. Súbory BAM využívajú formát pomenovania `SampleName_S#.bam`, kde # je číslo vzorky určené podľa poradia, v ktorom sú vzorky uvedené v rámci chodu. V režime s viacerými uzlami je S# nastavené na S1 bez ohľadu na poradie vzorky.

Súbory BAM obsahujú časť s hlavičkou a časť so zarovnaním:

- Hlavička – Obsahuje informácie o celom súbore (napr. názov vzorky, dĺžka vzorky a metóda zarovnaní). Zarovnaní v časti so zarovnaním sú spojené so špecifickými informáciami v časti s hlavičkou.
- Zarovnaní – Obsahuje názov čítania, sekvenciu čítania, informácie o zarovnaní a prispôbené príznaky. Názov čítania zahŕňa chromozóm, štartovacie súradnice, kvalitu zarovnaní a reťazec deskriptora zhody.

V časti so zarovnaním sú uvedené nasledujúce informácie pre každé čítanie alebo čítaný pár:

- AS: kvalita zarovnaní na oboch koncoch.
- RG: skupina čítania, ktorá indikuje počet čítaní špecifickej vzorky.
- BC: štítok (tzv. tag) čiarového kódu, ktorý indikuje ID demultiplexovanej vzorky súvisiacej s čítaním.
- SM: kvalita zarovnaní na jednom konci.
- XC: reťazec deskriptora zhody.

- XN: Značka názvu Amplicon, ktorá zaznamenáva ID amplikónu spojené so súbormi indexu čítania BAM (*.bam.bai) poskytuje index príslušného súboru BAM.

Súbory VCF

Súbory analýzy variantu (*.vcf) obsahujú informácie o nájdených variantoch na špecifických pozíciách v referenčnom genóme.

V hlavičke súborov VCF sa uvádza verzia formátu súboru VCF a verzia analýzy (tzv. caller) variantu a tiež poznámky použité vo zvyšku súboru. V hlavičke súboru VCF je tiež uvedený súbor referenčného genómu a súbor BAM. V poslednom riadku hlavičky sú uvedené nadpisy stĺpcov pre riadky s údajmi. V každom riadku s údajmi súborov VCF sú uvedené informácie o jednom variante.

Tabuľka 1 Hlavičky súborov VCF

Hlavička	Popis
CHROM	Chromozóm referenčného genómu. Chromozómy sa zobrazujú v rovnakom poradí ako v referenčnom súbore FASTA.
POS	Jednobázová pozícia variantu v referenčnom chromozóme. Pre varianty s jedným nukleotidom (SNV) je táto pozícia referenčnou bázou s variantom. V prípade indelov je táto pozícia referenčnou bázou bezprostredne pred variantom.
ID	Číslo rs (referenčné SNP) pre SNP získané z dbSNP.txt (ak je to uplatniteľné). Ak sa na tomto mieste nachádza viacero čísel rs, zoznam je oddelený bodkočiarkami. Ak pre túto pozíciu neexistuje žiadny dbSNP vstup, použije sa značka chýbajúcej hodnoty („“).
REF	Referenčný genotyp. Napríklad delécia jedného T je vyjadrená ako referenčný TT alebo alternatívny T. Variant s jedným nukleotidom z A na T je vyjadrený ako referenčný A a alternatívny T.
ALT	Alely, ktoré sa líšia od referenčného čítania. Napríklad inzercia jediného T je vyjadrená ako referenčný A a alternatívny AT. Variant s jedným nukleotidom z A na T je vyjadrený ako referenčný A a alternatívny T.
QUAL	Skóre kvality podľa Phredovej škály priradené k prekladaču (tzv. caller) variantu. Vyššie skóre označuje vyššiu istotu vo variante a nižšiu pravdepodobnosť chýb. Pre skóre kvality Q je odhadovaná pravdepodobnosť chyby $10^{-Q/10}$. Napríklad súbor stanovení Q30 má mieru výskytu chyby 0,1 %. Mnoho prekladačov variantov prideliuje skóre kvality na základe svojich štatistických modelov, ktoré sú vysoké v porovnaní s pozorovanou mierou výskytu chýb.

Tabuľka 2 Poznámky v súboroch VCF

Hlavička	Popis
FILTER	<p>Ak je vyhovené všetkým filtrom, do stĺpca na filter sa zapíše PASS (Úspešné). Možné pracovné postupy FILTER Germline zahŕňajú:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL – použije sa, ak skóre QUAL variantu SNP nespĺňa prahovú hodnotu • DRAGENIndelHardQUAL – použije sa, ak skóre QUAL indel variantu nespĺňa prahovú hodnotu • LowDepth – filtrované pracoviskom, pretože hĺbka pokrytia nespĺňa prahovú hodnotu • LowGQ – filtrované pracoviskom, pretože kvalita genotypu nespĺňa prahovú hodnotu • PloidyConflict – analýza genotypu od tzv. callera variantu nie je v súlade s ploidiou chromozómov • base_quality – filtrované pracoviskom, pretože stredná kvalita bázy alt čítaní na tomto mieste nespĺňa prahovú hodnotu • filtered_reads – filtrované pracoviskom, pretože bola odfiltrovaná príliš veľká časť čítaní • fragment_length – filtrované pracoviskom, pretože absolútny rozdiel medzi strednou dĺžkou fragmentu alt čítaniami a strednou dĺžkou fragmentu ref čítaní na tomto mieste prekračuje prahovú hodnotu • low_depth – filtrované pracoviskom, pretože hĺbka čítania je príliš nízka • low_frac_info_reads – filtrované pracoviskom, pretože zlomok informatívnych čítaní je pod prahovou hodnotou • low_normal_depth – filtrované pracoviskom, pretože normálna hĺbka čítania vzorky je príliš nízka • long_indel – filtrované pracoviskom, pretože indel je príliš dlhý • mapping_quality – filtrované pracoviskom, pretože stredná kvalita mapovania alt čítaní na tomto mieste nespĺňa prahovú hodnotu • multiallelic – filtrované pracoviskom, pretože viac ako dve alt alely prechádzajú LOD nátoru • non_homref_normal – filtrované pracoviskom, pretože bežný genotyp vzorky nie je homozygotný referenčný • no_reliable_supporting_read – filtrované pracoviskom, pretože neexistuje spoľahlivé podporné somatické čítanie • panel_of_normals – nachádza sa aspoň v jednej vzorke v paneli bežných súborov vcf • read_position – filtrované pracoviskom, pretože medián vzdialeností medzi začiatkom/koncom čítania a týmto miestom je pod prahovou hodnotou • RMxNRepeatRegion – filtrované pracoviskom, pretože celá alebo časť alely variantu je opakovaním referenčnej

Hlavička	Popis
FILTER (pokračovanie)	<ul style="list-style-type: none"> • strand_artifact – filtrované pracoviskom z dôvodu závažného predpätia vlákna • str_contraction – filtrované pracoviskom z dôvodu podozrenia na chybu PCR, keď je alela alt o jednu opakovaciu jednotku menšia ako referenčná • too_few_supporting_reads – filtrované pracoviskom, pretože vo vzorke nádoru je príliš málo podporných čítaní • weak_evidence – skóre somatických variantov nespĺňa prahovú hodnotu <p>Možné položky FILTRA pre somatický pracovný postup zahŕňajú:</p> <ul style="list-style-type: none"> • base_quality – filtrované pracoviskom, pretože stredná kvalita bázy alt čítaní na tomto mieste nespĺňa prahovú hodnotu • filtered_reads – filtrované pracoviskom, pretože bola odfiltrovaná príliš veľká časť čítaní • fragment_length – filtrované pracoviskom, pretože absolútny rozdiel medzi strednou dĺžkou fragmentu alt čítaniami a strednou dĺžkou fragmentu ref čítaní na tomto mieste prekračuje prahovú hodnotu • low_depth – filtrované pracoviskom, pretože hĺbka čítania je príliš nízka • low_frac_info_reads – filtrované pracoviskom, pretože zlomok informatívnych čítaní je pod prahovou hodnotou • low_normal_depth – filtrované pracoviskom, pretože normálna hĺbka čítania vzorky je príliš nízka • long_indel – filtrované pracoviskom, pretože indel je príliš dlhý • mapping_quality – filtrované pracoviskom, pretože stredná kvalita mapovania alt čítaní na tomto mieste nespĺňa prahovú hodnotu • multiallelic – filtrované pracoviskom, pretože viac ako dve alt alely prechádzajú LOD nádoru • non_homref_normal – filtrované pracoviskom, pretože bežný genotyp vzorky nie je homozygotný referenčný • no_reliable_supporting_read – filtrované pracoviskom, pretože neexistuje spoľahlivé podporné somatické čítanie • panel_of_normals – nachádza sa aspoň v jednej vzorke v paneli bežných súborov vcf • read_position – filtrované pracoviskom, pretože medián vzdialeností medzi začiatkom/koncom čítania a týmto miestom je pod prahovou hodnotou • RMxNRepeatRegion – filtrované pracoviskom, pretože celá alebo časť alely variantu je opakovaním referenčnej • strand_artifact – filtrované pracoviskom z dôvodu závažného predpätia vlákna • str_contraction – filtrované pracoviskom z dôvodu podozrenia na chybu PCR, keď je alela alt o jednu opakovaciu jednotku menšia ako referenčná • too_few_supporting_reads – filtrované pracoviskom, pretože vo vzorke nádoru je príliš málo podporných čítaní

Hlavička	Popis
INFO (INFORMÁCIE)	<ul style="list-style-type: none"> • weak_evidence – skóre somatických variantov nespĺňa prahovú hodnotu • systematic_noise – filtrované pracoviskom na základe dôkazov o systematickom šumu v bežných hodnotách <p>Medzi možné záznamy o INFORMÁCIÁCH z pracovného postupu Germline patria:</p> <ul style="list-style-type: none"> • AC – počet alel v genotypoch pre každú ALT alelu v tom istom poradí, ako sú uvedené. • AF – frekvencia alel pre každú ALT alelu v takom istom poradí, ako sú uvedené. • AN – celkový počet alel v stanovených genotypoch. • DB – dbSNP členstvo. • FS – P-hodnota vyhodnotená pomocou Fisherovho presného testu na zistenie vychýlenia vlákna. • QD – spoľahlivosť/kvalita variantov podľa hĺbky. • R2_5P_bias – skóre založené na spárovanom vychýlení a vzdialenosti od 5 hlavných koncov. • SOR – symetrický pomer pravdepodobnosti 2 x 2 pohotovostnej tabuľky na zistenie vychýlenia vlákna. • DP – približná hĺbka čítania (informatívna a neinformatívna); niektoré čítania mohli byť filtrované na základe mapq atď. • END – poloha zastavenia intervalu. • FractionInformativeReads – podiel informatívnych čítaní z celkových čítaní. • MQ – kvalita mapovania RMS. • MQRankSum – Z-skóre z Wilcoxonovho testu súčtu poradia kvality mapovania Alt vs. Ref. • ReadPosRankSum – Z-skóre z Wilcoxonovho testu súčtu poradia vychýlenia pozície Alt vs. Ref. • SOMATIC – aspoň jeden variant v tejto pozícii je somatický. <p>Medzi možné zadania INFORMÁCIÍ o pracovných postupoch Somatic patria:</p> <ul style="list-style-type: none"> • DP – približná hĺbka čítania (informatívna a neinformatívna); niektoré čítania mohli byť filtrované na základe mapq atď. • END – poloha zastavenia intervalu. • FractionInformativeReads – podiel informatívnych čítaní z celkových čítaní. • MQ – kvalita mapovania RMS. • MQRankSum – Z-skóre z Wilcoxonovho testu súčtu poradia kvality mapovania Alt vs. Ref. • ReadPosRankSum – Z-skóre z Wilcoxonovho testu súčtu poradia vychýlenia pozície Alt vs. Ref. • AQ – systematické skóre šumu.

Hlavička	Popis
FORMAT (FORMÁT)	<ul style="list-style-type: none"> • hotspot – známa somatická lokalita, ktorá sa používa na zvýšenie dôvery v analýzu. • SOMATIC – aspoň jeden variant v tejto pozícii je somatický. <p>V stĺpci formátu sú políčka oddelené dvojbodkami. Napríklad GT:GQ.</p> <p>Dostupné polia pracovného postupu Germline zahŕňajú:</p> <ul style="list-style-type: none"> • AD – alelické hĺbky (počítanie iba informatívnych čítaní z celkových čítaní) pre ref. a alt. alely v uvedenom poradí. • AF – alelové frakcie pre alt alely v uvedenom poradí. • DP – približná hĺbka čítania (čítanie s MQ = 255 alebo zlými väzbami sú odfiltrované). • F1R2 – počet čítaní v párovej orientácii F1R2 podporujúcich každú alelu. • F2R1 – počet čítaní v párovej orientácii F2R1 podporujúcich každú alelu. • GP – posteriórne pravdepodobnosti podľa Phred skóre genotypov definované v špecifikácii VCF. • GQ – kvalita genotypu. • GT – genotyp. 0 zodpovedá referenčnej báze, 1 zodpovedá prvému záznamu v stĺpci ALT atď. Lomka (/) znamená, že k dispozícii nie sú žiadne informácie o fázovaní. • MB – štatistiky jednotlivých komponentov na každú vzorku na detekciu vychýlenia párovania. • PL – normalizované, pravdepodobnosti pre genotypy podľa Phred skóre, ako sú definované v špecifikácii VCF. • PRI – predchádzajúce pravdepodobnosti pre genotypy podľa Phred skóre. • PS – fyzikálne informácie o ID fázovania, kde každé jedinečné ID v rámci danej vzorky (ale nie v rámci vzoriek) spája záznamy v rámci skupiny fázovania. • SB – štatistiky jednotlivých vzoriek, ktoré obsahujú Fisherov presný test na detekciu vychýlenia vlákna. • SQ – somatická kvalita. <p>Dostupné polia pracovného postupu Somatic zahŕňajú:</p> <ul style="list-style-type: none"> • AD – alelické hĺbky (počítanie iba informatívnych čítaní z celkových čítaní) pre ref. a alt. alely v uvedenom poradí. • AF – alelové frakcie pre alt alely v uvedenom poradí. • DP – približná hĺbka čítania (čítanie s MQ = 255 alebo zlými väzbami sú odfiltrované). • F1R2 – počet čítaní v párovej orientácii F1R2 podporujúcich každú alelu. • F2R1 – počet čítaní v párovej orientácii F2R1 podporujúcich každú alelu. • GT – genotyp. 0 zodpovedá referenčnej báze, 1 zodpovedá prvému záznamu v stĺpci ALT atď. Lomka (/) znamená, že k dispozícii nie sú žiadne informácie o fázovaní. • MB – štatistiky jednotlivých komponentov na každú vzorku na detekciu vychýlenia párovania.

Hlavička	Popis
SAMPLE (VZORKA)	<ul style="list-style-type: none"> • PS – fyzikálne informácie o ID fázovania, kde každé jedinečné ID v rámci danej vzorky (ale nie v rámci vzoriek) spája záznamy v rámci skupiny fázovania. • SB – štatistiky jednotlivých vzoriek, ktoré obsahujú Fisherov presný test na detekciu vychýlenia vlákna. • SQ – somatická kvalita. Stípec vzorky udáva hodnoty špecifikované v stípci FORMAT (FORMÁT).

Súbory VCF genómu

Súbory VCF (*.gvcf.gz) dodržiavajú súbor konvencií na vyjadrenie všetkých miest v rámci genómu v primerane kompaktnom formáte. Súbory gVCF obsahujú všetky miesta v oblasti záujmu v jednom súbore pre každú vzorku. Súbor gVCF zobrazuje absenciu stanovenia v pozíciách, ktoré nespĺnili všetky požiadavky filtrov. Štítok genotypu (GT) ./. označuje žiadne stanovenie.

Zobrazenie výsledkov analýzy

Aktuálne prebiehajúce chody sa zobrazia na karte Active (Aktívne). Dokončené chody sa zobrazia na karte Completed (Dokončené). Ďalšie informácie o zobrazovaní výsledkov nájdete v časti [Produktová dokumentácia NovaSeq 6000Dx \(dokument č. 200010105\)](#).

Technická pomoc

Technickú pomoc vám poskytne technická podpora spoločnosti Illumina.

Webová lokalita: www.illumina.com

E-mail: techsupport@illumina.com

Illumina Telefónne čísla technickej podpory

Región	Bezplatné	Medzinárodné
Austrália	+61 1800 775 688	
Rakúsko	+43 800 006249	+43 1 9286540
Belgicko	+32 800 77 160	+32 3 400 29 73
Kanada	+1 800 809 4566	
Čína		+86 400 066 5835
Dánsko	+45 80 82 01 83	+45 89 87 11 56
Fínsko	+358 800 918 363	+358 9 7479 0110
Francúzsko	+33 8 05 10 21 93	+33 1 70 77 04 46
Nemecko	+49 800 101 4940	+49 89 3803 5677
Hongkong, Čína	+852 800 960 230	
India	+91 8006500375	
Indonézia		0078036510048
Írsko	+353 1800 936608	+353 1 695 0506
Taliansko	+39 800 985513	+39 236003759
Japonsko	+81 0800 111 5011	
Malajzia	+60 1800 80 6789	
Holandsko	+31 800 022 2493	+31 20 713 2960
Nový Zéland	+64 800 451 650	
Nórsko	+47 800 16 836	+47 21 93 96 93
Filipíny	+63 180016510798	
Singapur	1 800 5792 745	
Južná Kórea	+82 80 234 5300	
Španielsko	+34 800 300 143	+34 911 899 417

Región	Bezplatné	Medzinárodné
Švédsko	+46 2 00883979	+46 8 50619671
Švajčiarsko	+41 800 200 442	+41 56 580 00 00
Taiwan, Čína	+886 8 06651752	
Thajsko	+66 1800 011 304	
Spojené kráľovstvo	+44 800 012 6019	+44 20 7305 7197
Spojené štáty americké	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

Karty bezpečnostných údajov (Safety data sheets, SDS) – k dispozícii na webovej lokalite spoločnosti Illumina na stránke support.illumina.com/sds.html.

Produktová dokumentácia – k dispozícii na stiahnutie z webovej lokality support.illumina.com.



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

CE



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

Austrálsky zadávateľ

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

NA DIAGNOSTICKÉ ÚČELY IN VITRO

© 2022 Illumina, Inc. Všetky práva vyhradené.

illumina[®]