

DRAGEN for Illumina DNA Prep with Enrichment Dx

Termékdokumentáció a NovaSeq 6000Dx készülékhez

Módosítási előzmények

Dokumentum	Dátum	Módosítások leírása
200014776 v02	2022. szeptember	A szövegről (*.txt) BED-re (*.bed) történő jegyzékfájl-formátum kijavítás a futtatás létrehozására vonatkozó utasításokban. A konszenzusos VCF-fájloknak az elemzés kimeneti részében lévő VCF-fájlokra történő javítása.
200014776 v01	2022. augusztus	Hozzáadva: Beállítások rész. Szisztematikus zajszűrés szakasz. A futtatás-létrehozási utasítások frissítése, a részletes leírás hozzáadása miatt. Elírások és nyelvtani hibák kijavítása. Annak meghatározása, hogy NovaSeq 6000Dx Instrument készülékkel való használat esetén az utasítások az alkalmazás célját szolgálják. A VCF kimeneti fájl tartalmáról szóló információk frissítése.
200014776 v00	2022. március	Első kiadás.

A jelen dokumentum és annak tartalma az Illumina, Inc. és annak leányvállalatai („Illumina”) tulajdonát képezi, és kizárólag a jelen dokumentumban ismertetett termék(ek) szerződésszerű működtetéséhez használható. Egyéb célokra nem használható. A dokumentum és annak tartalma az Illumina előzetes írásos engedélye nélkül ettől eltérő célokra nem használható és forgalmazható, továbbá semmilyen formában nem kommunikálható, hozható nyilvánosságra vagy reprodukálható. Az Illumina a jelen dokumentummal nem biztosít licencet a termék vásárlójának a harmadik felek szabadalmi, védjegyjogi, szerzői jogi, szokásjogi vagy egyéb oltalom alatt álló jogosultságaihoz.

A jelen dokumentumban szereplő utasításokat a kvalifikált és megfelelően képzett személyzetnek szigorúan be kell tartania az itt ismertetett termék(ek) megfelelő és biztonságos használata érdekében. A termék(ek) használata előtt a felhasználó köteles átolvasni és értelmezni a jelen dokumentumban leírtakat.

AZ ITT SZEREPLŐ INFORMÁCIÓK ELOLVASÁSÁNAK VAGY AZ UTASÍTÁSOK BETARTÁSÁNAK ELMULASZTÁSA ESETÉN A TERMÉK(EK) MEGSÉRÜLHETNEK, ILLETVE SZEMÉLYI SÉRÜLÉS KÖVETKEZHET BE, IDEÉRTVE A FELHASZNÁLÓKAT ÉS MÁSOKAT IS, ILLETVE EGYÉB ANYAGI KÁROK KÖVETKEZHETNEK BE. EZENFELÜL ILYEN ESETEKBE A TERMÉK (EK)RE VONATKOZÓ GARANCIA ÉRVÉNYÉT VESZTI.

AZ ILLUMINA SEMMIFÉLE FELELŐSSÉGET NEM VÁLLAL AZ ITT BEMUTATOTT TERMÉK(EK) HELYTELEN HASZNÁLATÁBÓL FAKADÓ KÁROKÉRT (AZ ALKATRÉSZEKET ÉS A SZOFTVERT IS IDEÉRTVE).

© 2022 Illumina, Inc. Minden jog fenntartva.

Minden védjegy az Illumina, Inc., illetve az adott tulajdonosok tulajdonát képezi. A márkanevekkel kapcsolatos információkat lásd a www.illumina.com/company/legal.html internetes oldalon.

Tartalomjegyzék

Módosítási előzmények	ii
Áttekintés	1
Elemzési módszerek	1
Futtatás létrehozása	4
Beállítások	5
Elemzési eredményt tartalmazó fájlok	7
FASTQ fájlok	8
BAM fájlok	8
VCF-fájlok	9
Elemzési eredmények megtekintése	14
Műszaki támogatás	15

Áttekintés

A DRAGEN™ for Illumina® DNA Prep with Enrichment Dx alkalmazás elvégzi a demultiplikálást, a FASTQ generálását, a kiolvasási leképezést, és a referenciagenomhoz való illesztést és a variánsazonosítást a kiválasztott elemzési munkafolyamattól függően.

Elemzési módszerek

A DRAGEN for Illumina DNA Prep with Enrichment Dx elvégzi a demultiplikálást, a FASTQ generálását, a leolvasási leképezést és a referenciagenomhoz való illesztést a kiválasztott munkafolyamattól függően:

- FASTQ létrehozása
- Csírvonal FASTQ és VCF generálás
- Szomatikus FASTQ és VCF generálás

FASTQ létrehozása

Az összeállított szekvenciák mintáknaként FASTQ-fájlokba vannak írva. A FASTQ-fájlok olyan szöveges fájlok, amelyek csak egy minta szekvenálási adatait és minőségi pontszámait tartalmazzák. Minden egyes mintához külön FASTQ-fájlok jönnek létre áramlási cella sávonként, szekvenálási leolvasásonként. A futtatás beállítása során megadott minta neve szerepel a FASTQ fájlnevben. A FASTQ-fájlok az igazítás elsődleges bemeneti adatai. A FASTQ generálás első lépése a demultiplikálás. A demultiplikálás szűrőn átmenő klasztereket rendel hozzá egy mintához, az egyes indexkiolvasási szekvenciákat összehasonlítva a futtatáshoz megadott index szekvenciákkal. Ebben a lépésben nem történik minőségi értékek figyelembevétele. Az indexkiolvasások azonosítása a következő lépésekben történik:

- A rendszer 1-től kezdődően számozza a mintákat, a futtatáshoz való felsorolás sorrendjében.
- A 0-s mintaszám a mintához hozzá nem rendelt klaszterek számára van fenntartva.
- A klaszterek akkor kerülnek hozzárendelésre egy mintához, ha az indexszekvencia pontosan megegyezik, vagy ha indexkiolvasásonként legfeljebb egy eltérés van.

A szoftver ORA-tömörítést ad hozzá a FASTQ-fájlok tömörítéséhez. ORA (*.ora) formátum használata esetén a veszteségmentes tömörítés biztosítása érdekében a FASTQ tartalom md5 ellenőrzőösszege a tömörítési és kibontási ciklus után megmarad.

DNS leképezés és igazítás

A leképezés első fázisa a magok generálása a leolvasásból, majd a pontos egyezések keresése a referencia genomban. Ezeket az eredményeket ezután a teljes Smith-Waterman igazítások futtatásával igazítjuk a legnagyobb sűrűségű megegyezéseket tartalmazó helyeken. Ez a jól dokumentált algoritmus úgy működik, hogy összehasonlítja az olvasás minden egyes pozícióját a referencia összes jelölt

pozíciójával. Ezek az összehasonlítások megfelelnek a leolvasás és a referencia közötti potenciális igazítások mátrixának. A Smith-Waterman minden egyes ilyen lehetséges igazítási pozícióhoz pontszámokat generál, amelyek annak értékelésére szolgálnak, hogy az adott mátrixcellán áthaladó legjobb igazítás nukleotidegyezéssel vagy eltéréssel (átlós mozgás), delécióval (vízszintes mozgás) vagy beillesztéssel (függőleges mozgás) éri-e el azt. Az olvasás és a hivatkozás közötti egyezés bónuszt jelent a pontszámánál, az eltérés vagy az indel pedig büntetést von maga után. Az összességében legmagasabb pontszámú útvonal a mátrixon keresztül a kiválasztott igazítás.

Az ebben az algoritmusban a pontszámokhoz kiválasztott specifikus értékek azt jelzik, hogy hogyan lehet kiegyensúlyozni – több lehetséges értelmezéssel bíró illeszkedés esetén – egy vagy több SNP-vel szemben az indel lehetőségét, vagy egy vágás nélküli illeszkedés preferenciáját. Az alapértelmezett DRAGEN pontozási értékek észszerűek a mérsékelt hosszúságú kiolvasások és egy teljes humán referenciagenom illesztéséhez a variánsazonosítási alkalmazásokban. A Smith-Waterman pontozási paraméterek bármely halmaza a genomikai mutáció és szekvenálási hibák pontatlan modelljét képviseli. A különböző módon beállított illesztéspontozási értékek egyes alkalmazásoknál megfelelőbbek lehetnek.

DRAGEN csírvonal variánsazonosítás

A DRAGEN csírvonal kis variánsazonosítója a feltérképezett és illesztett DNS-kiolvasásokat használja bemenetként, és az SNP-eket, valamint az indeleket oszloponkénti detektálás és a haplotípusok helyi *de novo* összeállításának kombinációjával azonosítja.

A lehívható hivatkozási tartományokat először megfelelő illesztési lefedettséggel kell azonosítani. Ezekben a hivatkozási tartományokon belül a rendszerezett kiolvasások gyors szkennelése azonosítja az aktív tartományokat, amelyek a variáns bizonyítékával rendelkező halmozott oszlopok közepe táján helyezkednek el. Az aktív tartományok elégséges kontextussal vannak kitöltve ahhoz, hogy a közelben lévő jelentős, nem referenciális tartalmakat lefedjék. Ha indelre utaló jelek vannak, az aktív tartományok további kitöltést kapnak.

Az egymáshoz igazított leolvasásokat az egyes aktív tartományokban levágják, és egy De Bruijn-gráfba illesztik. A levágott kiolvasások széleit a megfigyelési számok súlyozzák, a referenciaszekvencia pedig gerincként szolgál. Némi gráftisztítás és egyszerűsítés után az összes forrástól a fogadásig tartó útvonal jelölt haplotípusként kerül kivonásra. Minden egyes haplotípus Smith-Waterman-igazításra kerül a referenciagenomhoz, hogy azonosítani lehessen az általa képviselt variánsokat. Ez az eseménysorozat pozícióalapú észleléssel egészíthető ki. Minden egyes leolvasási haplotípus pár esetében a leolvasás megfigyelésének $P(r|H)$ valószínűségét – feltételezve, hogy a haplotípus a valódi kiindulási minta – egy pár rejtett Markov-modell (HMM) segítségével becsüljük meg.

Az aktív tartomány referenciapozíciója szerinti szkennelés során a lehetséges genotípusok a variáns események (SNP-k vagy indelek) diploid kombinációiból jönnek létre. Minden egyes eseménynél (beleértve a referenciát is) az egyes egymást átfedő kiolvasások megfigyelésének $P(r|e)$ feltételes valószínűségét az eseményt támogató haplotípusok maximális $P(r|H)$ valószínűségeként becsüljük. Ezeket egy genotípusra (eseménypárra) vonatkozó $P(r|e1e2)$ feltételes valószínűséggé kombináljuk, és

megszorozva megkapjuk a teljes olvasási halmaz megfigyelésének $P(R|e1e2)$ feltételes valószínűségét. A Bayes-képlet segítségével kiszámítjuk az egyes diploid genotípusok $P(e1e2|R)$ utólagos valószínűségét, és kijelöljük a nyertest.

A mérhető többmintás variánsazonosításhoz használt gVCF módban a DRAGEN csírvonal kis variánsazonosítója mintánként futtatható közbelső genomikai variánsazonosító fájl (gVCF) létrehozásához. A gVCF ezután felhasználható több minta hatékony közös genotipizálására, amely lehetővé teszi a minták gyors növekményes feldolgozását és a nagy kohorszméretekre történő skálázást.

Mivel a DRAGEN csírvonal kis variánsazonosítója rendelkezik a korrelált hibáknak a valódi variánsoktól való hatékony megkülönböztetésére alkalmas algoritmusokkal, a szűrési szabályok nagyon egyszerűek.

DRAGEN szomatikus variánsazonosítás

A DRAGEN Somatic Small Variant Caller szomatikus kis változatú variánsazonosítója bemenetként veszi a leképezett és illesztett DNS-leolvasásokat, és az aktív tartományban lévő haplotípusok helyi *de novo* összeállításával azonosítja az SNV-ket és indeleket.

A lehívható hivatkozási tartományokat először megfelelő illesztési lefedettséggel kell azonosítani. Ezekben a hivatkozási tartományokon belül a rendszerezett kiolvasások szkennelése azonosítja az aktív tartományokat, amelyek a tumorleolvasásokban egy variánsra utaló bizonyítékkal rendelkező halmozott oszlopok középpontja környékén helyezkednek el. Az aktív tartományok elégséges kontextussal vannak kitöltve ahhoz, hogy a közelben lévő jelentős, nem referenciális tartalmakat lefedjék. Ha indelre utaló jelek vannak, az aktív tartományok további kitöltést kapnak.

Az egymáshoz igazított leolvasásokat az egyes aktív tartományokban levágják, és egy De Bruijn-gráfba illesztik. A levágott kiolvasások széleit a megfigyelési számok súlyozzák, a referenciaszekvencia pedig gerincként szolgál. Némi gráftisztítás és egyszerűsítés után az összes forrástól a fogadásig tartó útvonal jelölt haplotípusként kerül kivonásra. Minden egyes haplotípus Smith-Waterman-igazításra kerül a referenciagenomhoz, hogy azonosítani lehessen az általa képviselt variánsokat. Minden egyes kiolvasási haplotípus pár esetében a kiolvasás megfigyelésének valószínűségét $P(r|H)$ egy pár rejtett Markov-modell (HMM) segítségével becsüljük meg, feltételezve, hogy a haplotípus a valódi kiindulási minta.

A TLOD-pontszám meghatározásához a DRAGEN Somatic Small Variant Caller először referenciapozíciónként beszkenyel minden egyes jelölt szomatikus eseményt, valamint az aktív régióban lévő referenciaeseményt is. Az egyes egymást átfedő leolvasások megfigyelésének feltételes valószínűsége $P(r|e)$ maximumnak becsülhető $P(r|H)$ az eseményt alátámasztó haplotípusok esetében. Ezeket a $P(r|E)$ feltételes valószínűséggé kombináljuk egy E eseményhipotézisre, amely a referencia- és a jelölt szomatikus allél keverékét tartalmazza a lehetséges allélgyakoriságok tartományában, és megszorozzuk, hogy megkapjuk a teljes olvasási halmaz megfigyelésének $P(R|E)$ feltételes valószínűségét. Ebből számítják ki a TLOD-pontszámot annak bizonyítékaként, hogy egy adott lókuszon ALT-allél van jelen a tumormintában.

Futtatás létrehozása

A következő lépésekkel állíthatja be a futtatást az Illumina Run Manager szoftveren, vagy a NovaSeq 6000Dx segítségével, vagy egy hálózatba kapcsolt számítógépen a böngésző használatával. A mintaadatok manuálisan vagy egy mintalap importálásával adhatók meg.

Alkalmazás és futtatási beállítások

1. A Runs (Futtatások) képernyőn válassza a **Create Run** (Futtatás létrehozása) lehetőséget.
2. Válassza ki a DRAGEN for Illumina DNA Prep with Enrichment Dx alkalmazást, majd kattintson a **Next** (Következő) gombra.
3. A Run Settings (Futtatási beállítások) képernyőn adjon meg egy futtatásnevet. A futtatás neve azonosítja a futtatást a szekvenálástól az elemzésig.
4. **[Választható]** Adja meg a futtatás leírását a futtatás további azonosításához.
5. Győződjön meg arról, hogy a kiválasztott könyvtár-előkészítő készlet Illumina DNA Prep with Enrichment Dx könyvtár-előkészítő készlet.
6. Válassza ki a kívánt index adapterkészletét.
7. Adja meg a kiolvasás hosszát.
Az 1. és 2. kiolvasás alapértelmezett értéke 151 ciklus.
Az 1. és 2. index rögzített értéke 10 ciklus.
8. **[Választható]** Adjon meg egy könyvtárkémcső-azonosítót.
9. Válassza a **Next** (Tovább) lehetőséget.

Mintaadatok

A Sample Data (Mintaadatok) képernyőn lévő táblázat segítségével manuálisan vihet be mintaadatokat. Másik lehetőségként válassza az **Import Samples** (Minták importálása) lehetőséget a mintaadatok feltöltéséhez. A mintaadatok importálásával kapcsolatos információkért lásd [Minták importálása a\(z\) 5. oldalon](#) fejezetet.

Minták manuális bevitele

1. Adjon meg egy egyedi mintaazonosítót a Sample ID (Mintaazonosító) mezőben.
2. Az üregpozíció kiválasztásához használja a **Plate - Well Position** (Lemez – üregpozíció) elemet.
Az i7 Index, Index 1, i5 Index és Index 2 mezők automatikusan kitöltésre kerülnek.
3. **[Választható]** Adja meg a könyvtár nevét.
4. Adjon hozzá sorokat, és szükség szerint ismétlje meg az 1–3. lépéseket, amíg az összes mintát hozzá nem adta a táblázathoz.
5. Válassza a **Next** (Tovább) lehetőséget.

Minták importálása

Ha egy hálózatra kapcsolt számítógépen lévő böngésző segítségével tervez futtatást az Illumina Run Manager szoftverben, a Sample Data (Mintaadatok) képernyőn letölthető egy sablonfájl (*.csv).

1. Üres CSV-fájl letöltéséhez válassza a **Download Template** (Sablon letöltése) lehetőséget.
2. Adja meg a mintaadatokat a CSV-fájlban, és mentse el a fájlt.
A CSV-mintalapfájl a következő adatoszlopokat tartalmazza: Mintaazonosító, lemez – üregpozíció, **Opcionális** könyvtár neve.
3. Válassza az **Import Samples** (Minták importálása) lehetőséget, és tallózzon a CSV fájl helyére.
4. Válassza a **Next** (Tovább) lehetőséget.

Elemzési beállítások

1. Válassza ki a kívánt elemzési munkafolyamatot:
 - FASTQ létrehozása
 - Csírvonal FASTQ és VCF létrehozása csírvonal-munkafolyamathoz
 - Szomatikus FASTQ és VCF létrehozása szomatikus munkafolyamathoz
2. **[Opcionális]** Ha kívánja, jelölje be az **Generate ORA compressed FASTQs** (ORA tömörített FASTQ-k létrehozása) jelölőnégyzetet a FASTQ ORA tömörítés engedélyezéséhez.
3. **[VCF létrehozásának munkafolyamatai]** Használja a **Manifest File Selection** (Jegyzékfájl kiválasztása) legördülő menüt a jegyzékfájl kiválasztásához.
A jegyzékfájl bevitele szükséges a DRAGEN for Illumina DNA Prep with Enrichment Dx szoftverhez. A jegyzékfájl egy tabulátorral elválasztott BED (*.bed) fájl, amely meghatározza a célzott hivatkozási tartományok nevét és helyét.
4. **[Szomatikus FASTQ és VCF létrehozásának munkafolyamata]** A zajfájl kiválasztásához használja a **Noise File Selection** (Zajfájl kiválasztása) legördülő menüt.
A szisztematikus zaj kiszűréséhez megadható egy helyspecifikus zajszintű BED fájl. További információkért lásd a [Zajsűrés a\(z\) 6. oldalon](#) című részt.
5. Válassza a **Next** (Tovább) lehetőséget.

Futtatás Áttekintés

1. A Review (Áttekintés) képernyőn tekintse át a Run Settings (Futtatási beállítások), a Sample Data (Mintaadatok) és az Analysis Settings (Elemzési beállítások) képernyőkre bevitt adatokat.
2. Válassza a **Save** (Mentés) lehetőséget.
A futtatás mentésre kerül a Planned (Tervezett) lapon a Runs (Futtatások) képernyőn.

Beállítások

Válassza ki az alkalmazást az Applications (Alkalmazások) képernyőn az aktuális beállítások megtekintéséhez és a beállítások módosításához.

Konfiguráció

A konfigurációs képernyő a következő alkalmazásbeállításokat jeleníti meg:

- **Library Prep Kits** (Könyvtári előkészítő készletek)—Megjeleníti az alkalmazás alapértelmezett könyvtári előkészítő készletét. Ez a beállítás nem módosítható.
- **Index Adapter Kits** (Index adapterkészletek)—Megjeleníti az alkalmazáshoz tartozó alapértelmezett index adapterkészletet. Ez a beállítás nem módosítható.
- **Read lengths** (Beolvasások hossza)—A beolvasási hosszak alapértelmezés szerint 151-re vannak állítva az alkalmazáshoz, de a futtatás létrehozása során módosíthatók.
- **Manifest and Noise Files** (Jegyzék- és zajfájlok)—A jegyzék- és zajfájlok beállításainak feltöltése és módosítása.
 - Válassza az **Upload File** (Fájl feltöltése) lehetőséget a fájlok elemzésben való használatához.
 - Válassza a **Default** (Alapértelmezett) választógombot a fájl alapértelmezett jegyzék- vagy zajfájlként való beállításához a futtatás létrehozása során, amikor az alkalmazás ki van választva.
 - Jelölje be az **Enabled** (Engedélyezve) jelölőnégyzetet a futtatás létrehozása során a legördülő menüben megjelenítendő fájl beállításához.

Engedélyek

Használja a Permissions (Engedélyek) képernyőn található jelölőnégyzeteket az alkalmazás felhasználói hozzáféréseinek kezeléséhez.

Zajszűrés

A szomatikus munkafolyamat használata esetén szisztematikus zajszűrés áll rendelkezésre. A szűrő használható tumor-normál módban, de különösen hasznos olyan, csak tumoros futtatásoknál, ahol nem áll rendelkezésre párosított normál érték.

A szisztematikus zaj BED-et normál mintákból kell létrehozni. Javasoljuk, hogy olyan szisztematikus zaj-fájlokat hozzon létre, amelyek könyvtár-előkészítési, szekvenálási rendszer és panelspecifikusak. A zajfájl létrehozásához körülbelül 50 normál mintát ajánlott használni.

Elemzési eredményt tartalmazó fájlok

A DRAGEN for Illumina DNA Prep with Enrichment Dx a következő adatokat menti el az elemzési mappába. Csak a csírvonal és a szomatikus munkafolyamatok hoznak létre PDF-et.

- Használt jegyzékfájl
- Szoftververzió
- Mintaazonosítók
- Összes illesztett kiolvasás
- Illesztett kiolvasások százaléka mintánként
- Az azonosított SNV-k száma mintánként
- Az azonosított indelek száma mintánként
- Lefedettségi statisztika

Elemzés kimeneti fájlok

Az alkalmazás a következő kimeneti fájlokat hozza létre. A létrehozott pontos fájlok attól függenek, hogy melyik elemzési munkafolyamatot használják. A kimeneti fájlok az elemzési mappában találhatóak.

Kimeneti fájl	Leírás
FASTQ (*.fastq.gz vagy *.fastq.ora)	Köztes fájlok, amelyek minőségileg pontozott bázisazonosításokat tartalmaznak. A FASTQ-fájlok tartalmazzák az illesztési lépéshez használt elsődleges bemeneti adatokat. ORA-tömörítés kiválasztása esetén a fájlnev ezt tükrözi.
Illesztési BAM fájlok (*.bam)	Egy adott minta illesztett kiolvasásait tartalmazza.
Genom VCF fájlok (*.gvcf.gz)	Minden pozíció genotípusát tartalmazza, akár variánsként, akár referenciaként van azonosítva.
VCF fájlok (*.vcf.gz)	Az egyes pozíciókban azonosított variánsokat tartalmazza.
Futtatási mérőszámok jelentése (*.csv)	Minőségi mérőszámokat tartalmaz a futtatásról, beleértve a teljes hozamot és a Q30-pontszámot.

FASTQ fájlok

A FASTQ (*.fastq.gz, *.fastq.ora) egy szövegalapú fájlformátum, amely tartalmazza a kiolvasásonkénti bázisazonosításokat és minőségi értékeket. Minden fájl tartalmazza a következő adatokat:

- A mintaazonosító
- A szekvencia
- A plusz jel (+)
- A Phred minőségi pontszámok ASCII + 33 kódolt formátumban

A mintaazonosító a következőképpen van formázva.

```
@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=AAAAAAAAAAA9#:<#<;<<<????#=#
```

BAM fájlok

A BAM-fájl (*.bam) a SAM (szekvenálás-illesztési leképezés) fájl tömörített bináris verziója, amely legfeljebb 128 Mb-os illesztett szekvenciák megjelenítésére szolgál. A BAM fájlok a `SampleName_S#.bam` fájlnevezési formátumot használják. A # a minta száma, amelyet a minták futtatáshoz való listázásának sorrendje határoz meg. Multinode üzemmódban az S# S1-re van állítva, függetlenül a minta sorrendjétől.

A BAM-fájlok egy fejléct és egy illesztési szakaszt tartalmaznak:

- Fejléc—A teljes fájlra vonatkozó információkat tartalmaz, mint például a minta neve, a minta hossza és az illesztési módszer. Az illesztés szakaszban lévő illesztések a fejléc részben található specifikus információkhoz kapcsolódnak.
- Illesztések—Tartalmazza a kiolvasási nevet, a kiolvasási sorrendet, a kiolvasás minőségét, az illesztés adatait és az egyéni címkéket. A kiolvasási név tartalmazza a kromoszómát, az indítási koordinátát, az illesztés minőségét és az egyezés leíró karakterláncát.

Az illesztések rész az alábbi információkat tartalmazza minden kiolvasáshoz vagy kiolvasási párhoz:

- AS: Páros végű illesztési minőség.
- RG: Kiolvasási csoport, amely egy adott minta kiolvasásainak számát jelzi.
- BC: Vonalkódos címke, amely a kiolvasáshoz társított csatornabontott mintaazonosítót jelzi.
- SM: Páratlan végű illesztési minőség.

- XC: Megfelelést leíró karakterlánc.
- XN: A leolvasott BAM indexfájlokhoz (*.bam.bai) társított ampliton névcímke, amely rögzíti az ampliton azonosítót, biztosítja a megfelelő BAM fájl indexét.

VCF-fájlok

A variánsazonosítás formátumú (*.vcf) fájlok a referenciagenom meghatározott helyein található variánsokról tartalmaznak adatokat.

A VCF-fájl fejléce tartalmazza a VCF-fájlformátum verzióját és a variánsazonosító verzióját, és felsorolja a fájl fennmaradó részében használt annotációkat. A VCF-fejléc emellett tartalmazza a referenciagenom-fájlt és a BAM-fájlt is. A fejléc utolsó sora tartalmazza az adatsorok oszlopfejléceit. A VCF-fájl mindegyik adatsora egyetlen variánsra vonatkozó információkat tartalmaz.

1 táblázat VCF-fájl címsora

Címsor	Leírás
CHROM	A referenciagenom kromoszómája. A kromoszómák ugyanabban a sorrendben jelennek meg, mint a referencia FASTA-fájl.
POS	A variáns egybázisú helyzete a referencia kromoszómában. Egynukleotid-variánsok (SNV) esetén ez a pozíció a variáns referenciabázisa. Indelek esetén ez a pozíció a közvetlenül a variánst megelőző referenciabázis.
Azonosító	Az rs (referencia SNP) szám az SNP-re vonatkozóan, a <code>dbSNP.txt</code> -ből nyerve, ha van ilyen. Ha több rs-szám létezik ezen a helyen, a listát pontosvesszők választják el. Ha ezen a helyen nem létezik dbSNP-bejegyzés, akkor egy hiányzó értékjelölőt ('.') kell használni.
REF	A referencia genotípus. Például egyetlen T deléciója TT referenciaként és T alternatívaként jelenik meg. Az A-T egynukleotid-variáns A referenciaként és T alternatívaként jelenik meg.
ALT	A referencia kiolvasástól eltérő allélok. Például egy szimpla T inzerciója A referenciaként és AT alternatívaként jelenik meg. Az A-T egynukleotid-variáns A referenciaként és T alternatívaként jelenik meg.
QUAL	A variánsazonosítás által hozzárendelt Phred-skála szerinti minőségi pontszám. A magasabb pontszámok a variáns nagyobb megbízhatóságát jelzik, és a hibák kisebb valószínűségét. Q minőségi pontszám esetén a hiba becsült valószínűsége $10^{-(Q/10)}$. Például a Q30-as azonosítások 0,1%-os hibaarányal rendelkeznek. Számos variánsazonosító a statisztikai modelljei alapján határozza meg a minőségi pontszámokat, amelyek magasabbak a megfigyelt hibaarányhoz képest.

2 táblázat VCF-fájlfeliratok

Címsor	Leírás
SZŰRŐ	<p>Ha az összes szűrő sikeres, a PASS (Sikeres) felirat jelenik meg a szűrőoszlopban.</p> <p>A csírvonal-munkafolyamat lehetséges SZŰRŐ bejegyzései a következők:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL—Akkor alkalmazandó, ha az SNP variáns QUAL pontszáma nem éri el a küszöbértéket • DRAGENdelHardQUAL—Akkor alkalmazandó, ha az indel variáns QUAL pontszáma nem éri el a küszöbértéket • LowDepth—Helyszín szűrve, mert a lefedettség mélysége nem éri el a küszöbértéket • LowGQ—Helyszín szűrve, mert a genotípus minősége nem éri el a küszöbértéket • PloidyConflict—A variánsazonosítótól érkező genotípus-azonosítás nem felel meg a kromoszóma ploidának • base_quality—Helyszín szűrve, mert az alt kiolvasások medián bázisminősége ezen a lókuszon nem éri el a küszöbértéket • filtered_reads—Helyszín szűrve, mert a kiolvasások túl nagy töredéke került kiszűrésre • fragment_length—Helyszín szűrve, mert az alt kiolvasások medián fragmentumhossza és a ref kiolvasások medián fragmentumhossza közötti abszolút különbség ezen a lókuszon meghaladja a küszöbértéket • low_depth—Helyszín szűrve, mert a kiolvasási mélység túl alacsony • low_frac_info_reads—Helyszín szűrve, mert az informatív kiolvasások frakciója a küszöbérték alatt van • low_normal_depth—Helyszín szűrve, mert a normál mintakiolvasási mélység túl alacsony • long_indel—Helyszín szűrve, mert az indelhossz túl hosszú • mapping_quality—Helyszín szűrve, mert az alt kiolvasások medián leképezési minősége ezen a lókuszon nem éri el a küszöbértéket • multiallelic—Helyszín szűrve, mivel több mint két alt allél felel meg a tumor LOD-jának • non_homref_normal—Helyszín szűrve, mert a normál minta genotípusa nem homozigóta referencia • no_reliable_supporting_read—Helyszín szűrve, mert nincs megbízható támogató szomatikus kiolvasás • panel_of_normals—A normálértékek vcf paneljében legalább egy mintában látható • read_position—Helyszín szűrve, mert a kiolvasás kezdése/leállítása és a lókusztávolságok mediánja a küszöbérték alatt van • RMxNRepeatRegion—Helyszín szűrve, mert a variáns allél egésze vagy része a referencia ismétlése

Címsor	Leírás
SZŰRŐ (folytatás)	<ul style="list-style-type: none"> • strand_artifact—Helyszín szűrve, súlyos száleltérés miatt • str_contraction—Helyszín szűrve PCR hiba gyanúja miatt, amikor az alt allél egy ismétlési egységgel kevesebb, mint a referencia • too_few_supporting_reads—Helyszín szűrve, mert túl kevés támogató kiolvasás van a tumormintában • weak_evidence—A szomatikus variáns pontszáma nem éri el a küszöbértéket <p>A szomatikus munkafolyamat lehetséges SZŰRŐ bejegyzései a következők:</p> <ul style="list-style-type: none"> • base_quality—Helyszín szűrve, mert az alt kiolvasások medián bázisminősége ezen a lókuszon nem éri el a küszöbértéket • filtered_reads—Helyszín szűrve, mert a kiolvasások túl nagy töredéke lett kiszűrve • fragment_length—Helyszín szűrve, mert az alt kiolvasások medián fragmentumhossza és a ref kiolvasások medián fragmentumhossza közötti abszolút különbség ezen a lókuszon meghaladja a küszöbértéket • low_depth—Helyszín szűrve, mert a kiolvasási mélység túl alacsony • low_frac_info_reads—Helyszín szűrve, mert az informatív kiolvasások frakciója a küszöbérték alatt van • low_normal_depth—Helyszín szűrve, mert a normál mintakiolvasási mélység túl alacsony • long_indel—Helyszín szűrve, mert az indelhossz túl hosszú • mapping_quality—Helyszín szűrve, mert az alt kiolvasások medián leképezési minősége ezen a lókuszon nem éri el a küszöbértéket • multiallelic—Helyszín szűrve, mivel több mint két alt allél felel meg a tumor LOD-jának • non_homref_normal—Helyszín szűrve, mert a normál minta genotípusa nem homozigóta referencia • no_reliable_supporting_read—Helyszín szűrve, mert nincs megbízható támogató szomatikus kiolvasás • panel_of_normals—A normálértékek vcf paneljében legalább egy mintában látható • read_position—Helyszín szűrve, mert a kiolvasás kezdése/leállítás és a lókusz közötti távolságok mediánja a küszöbérték alatt van • RMxNRepeatRegion—Helyszín szűrve, mert a variáns allél egésze vagy része a referencia ismétlése • strand_artifact—Helyszín szűrve, súlyos száleltérés miatt • str_contraction—Helyszín szűrve PCR hiba gyanúja miatt, amikor az alt allél egy ismétlési egységgel kevesebb, mint a referencia • too_few_supporting_reads—Helyszín szűrve, mert túl kevés támogató kiolvasás van a tumormintában • weak_evidence—A szomatikus variáns pontszáma nem éri el a küszöbértéket • systematic_noise—Helyszín szűrve, a normálértékekben lévő szisztematikus zajra utaló jelek alapján

Címsor	Leírás
INFO	<p>A csírvonal-munkafolyamat lehetséges INFO bejegyzései a következők:</p> <ul style="list-style-type: none"> • AC—Az allélok száma a genotípusokban minden egyes ALT allélnál, a felsorolt sorrendben. • AF—Minden egyes ALT allél allélgyakorisága, a felsorolt sorrendben. • AN—Az allélok teljes száma az azonosított genotípusokban. • DB—dbSNP tagság. • FS—Phred-skála szerinti p-érték a Fisher-féle egzakt teszt segítségével a száleltérés kimutatására. • QD—Variáns megbízhatóság/minőség mélység szerint. • R2_5P_bias—Pontszám, amely a páreltérésen és az 5 primer végtől mért távolságon alapul. • SOR—A 2x2-es kontingenciatáblázat szimmetrikus esélyhányadosa a száleltérések észlelésére. • DP—Hozzávetőleges kiolvasási mélység (informatív és nem informatív); egyes kiolvasások a mapq stb. alapján szűrve lehetnek. • END—Az intervallum befejezési pozíciója. • FractionInformativeReads—Az összes kiolvasásból megmutatkozó informatív kiolvasások hányada. • MQ—RSM leképezési minőség. • MQRankSum—Z-pontszám a Wilcoxon-rangsor összesítő tesztből, az Alt vs. Ref kiolvasás leképezési minőségei alapján. • ReadPosRankSum—Z-pontszám a Wilcoxon-rangsor összesítő tesztből, az Alt vs. Ref kiolvasás pozíciós eltérései alapján. • SOMATIC—Ebben a pozícióban legalább egy variáns szomatikus. <p>A szomatikus munkafolyamat lehetséges INFO bejegyzései a következők:</p> <ul style="list-style-type: none"> • DP—Hozzávetőleges kiolvasási mélység (informatív és nem informatív); egyes kiolvasások a mapq stb. alapján szűrve lehetnek. • END—Az intervallum befejezési pozíciója. • FractionInformativeReads—Az összes kiolvasásból megmutatkozó informatív kiolvasások hányada. • MQ—RSM leképezési minőség. • MQRankSum—Z-pontszám a Wilcoxon-rangsor összesítő tesztből, az Alt vs. Ref kiolvasás leképezési minőségei alapján. • ReadPosRankSum—Z-pontszám a Wilcoxon-rangsor összesítő tesztből, az Alt vs. Ref kiolvasás pozíciós eltérései alapján. • AQ—Szisztematikus zaj-érték. • hotspot—Ismert szomatikus hely, amely az azonosítás megbízhatóságának növelésére szolgál. • SOMATIC—Ebben a pozícióban legalább egy variáns szomatikus.

Címsor	Leírás
FORMÁTUM	<p>A formátum oszlop a kettőspontokkal elválasztott mezőket sorolja fel. Például GT:GQ.</p> <p>A csírvonal-munkafolyamat elérhető mezői a következők:</p> <ul style="list-style-type: none"> • AD—Allélmélységek (csak az információval bíró kiolvasások az összes kiolvasásból) a ref és alt allélokhoz a felsorolt sorrendben. • AF—Alt allélok allélfrakciói a felsorolt sorrendben. • DP—Hozzávetőleges kiolvasási mélység (az MQ = 255 értékű, vagy a rossz párral rendelkező kiolvasások szűrtek). • F1R2—Az egyes allélokat támogató, F1R2 párorientációjú kiolvasások száma. • F2R1—Az egyes allélokat támogató, F2R1 párorientációjú kiolvasások száma. • GP—A VCF specifikációban meghatározott genotípusok Phred-skálázott poszterior valószínűségei. • GQ—Genotípus minősége. • GT—Genotípus. A 0 a referenciabázisnak, az 1 az ALT oszlop első bejegyzésének felel meg, és így tovább. Az előre mutató perjel (/) azt jelzi, hogy nem áll rendelkezésre fázishatásra vonatkozó információ. • MB—Mintánkénti komponensstatisztikák a páreltérések észlelésére. • PL—A VCF specifikációban meghatározott genotípusok normalizált, Phred-skálázott valószínűsége. • PRI—A genotípusok Phred-skálázott korábbi valószínűsége. • PS—Fizikai fázisazonosító adat, ahol egy adott mintán belül (de nem a minták között) minden egyedi azonosító összekapcsolja a rekordokat egy fáziscsoporton belül. • SB—Mintánkénti komponensstatisztikák, amelyek a Fisher-féle egzakt tesztet tartalmazzák a száleltérés kimutatására. • SQ—Szomatikus minőség. <p>A szomatikus munkafolyamat elérhető mezői a következők:</p> <ul style="list-style-type: none"> • AD—Allélmélységek (csak az információval bíró kiolvasások az összes kiolvasásból) a ref és alt allélokhoz a felsorolt sorrendben. • AF—Alt allélok allélfrakciói a felsorolt sorrendben. • DP—Hozzávetőleges kiolvasási mélység (az MQ = 255 értékű, vagy a rossz párral rendelkező kiolvasások szűrtek). • F1R2—Az egyes allélokat támogató, F1R2 párorientációjú kiolvasások száma. • F2R1—Az egyes allélokat támogató, F2R1 párorientációjú kiolvasások száma. • GT—Genotípus. A 0 a referenciabázisnak, az 1 az ALT oszlop első bejegyzésének felel meg, és így tovább. Az előre mutató perjel (/) azt jelzi, hogy nem áll rendelkezésre fázishatásra vonatkozó információ. • MB—Mintánkénti komponensstatisztikák a páreltérések észlelésére. • PS—Fizikai fázisazonosító adat, ahol egy adott mintán belül (de nem a minták között) minden egyedi azonosító összekapcsolja a rekordokat egy fáziscsoporton belül.

Címsor	Leírás
FORMÁTUM	<ul style="list-style-type: none"> • SB—Mintánkénti komponensstatisztikák, amelyek a Fisher-féle egzakt tesztet tartalmazzák a száleltérés kimutatására. • SQ—Szomatikus minőség.
MINTA	A minta oszlop a FORMAT (Formátum) oszlopban meghatározott értékeket adja meg.

Genome VCF-fájlok

A Genome VCF (*.gvcf.gz) fájlok egy sor konvenciót követnek a genomon belüli összes hely észszerűen kompakt formátumban történő ábrázolására. A gVCF-fájlok minden egyes minta esetében egyetlen fájlban tartalmazzák a vizsgált tartományban található összes helyet. A gVCF fájl nem mutat azonosítást olyan pozíciókban, amelyek nem felelnek meg az összes szűrőnek. A ./ genotípus (GT) címke nem-azonosítást jelez.

Elemzési eredmények megtekintése

A jelenleg folyamatban lévő futtatások az Active (Aktív) lapon jelennek meg. A befejezett futtatások a Completed (Befejezett) lapon jelennek meg. Az eredmények megtekintésével kapcsolatos további információkért lásd: [NovaSeq 6000Dx Termékdokumentáció \(dokumentumszám: 200010105\)](#).

Műszaki támogatás

Ha műszaki támogatásra van szüksége, vegye fel a kapcsolatot az Illumina műszaki ügyfélszolgálattal.

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

Illumina műszaki ügyfélszolgálati telefonszámai

Régió	Ingyenesen hívható	Nemzetközi
Ausztrália	+61 1800 775 688	
Ausztria	+43 800 006249	+43 1 9286540
Belgium	+32 800 77 160	+32 3 400 29 73
Kanada	+1 800 809 4566	
Kína		+86 400 066 5835
Dánia	+45 80 82 01 83	+45 89 87 11 56
Finnország	+358 800 918 363	+358 9 7479 0110
Franciaország	+33 8 05 10 21 93	+33 1 70 77 04 46
Németország	+49 800 101 4940	+49 89 3803 5677
Hongkong, Kína	+852 800 960 230	
India	+91 8006500375	
Indonézia		0078036510048
Írország	+353 1800 936608	+353 1 695 0506
Olaszország	+39 800 985513	+39 236003759
Japán	+81 0800 111 5011	
Malajzia	+60 1800 80 6789	
Hollandia	+31 800 022 2493	+31 20 713 2960
Új-Zéland	+64 800 451 650	
Norvégia	+47 800 16 836	+47 21 93 96 93
Fülöp-szigetek	+63 180016510798	
Szingapúr	1 800 5792 745	
Dél-Korea	+82 80 234 5300	
Spanyolország	+34 800 300 143	+34 911 899 417

Régió	Ingyenesen hívható	Nemzetközi
Svédország	+46 2 00883979	+46 8 50619671
Svájc	+41 800 200 442	+41 56 580 00 00
Tajvan, Kína	+886 8 06651752	
Thaiföld	+66 1800 011 304	
Egyesült Királyság	+44 800 012 6019	+44 20 7305 7197
Egyesült Államok	+1 800 809 4566	+1 858 202 4566
Vietnám	+84 1206 5263	

Biztonsági adatlapok (SDS-ek)—Az Illumina support.illumina.com/sds.html címen elérhető weboldalán található.

Termék dokumentáció—Letölthető a support.illumina.com weboldalról.



Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (Észak-Amerikán kívül)
techsupport@illumina.com
www.illumina.com

CE



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

Ausztrál megbízó

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

IN VITRO DIAGNOSZTIKAI HASZNÁLATRA

© 2022 Illumina, Inc. Minden jog fenntartva.

illumina[®]