

DRAGEN za Illumina DNK Prep with Enrichment Dx na NextSeq 550Dx

Korisnički priručnik za aplikaciju

Ovaj dokument i njegov sadržaj vlasništvo su tvrtke Illumina, Inc. i njezinih povezanih društava („Illumina“) te su namijenjeni isključivo za ugovornu upotrebu klijentima u vezi s proizvodima opisanim u njemu. Dokument i njegov sadržaj ne smiju se upotrebljavati ni distribuirati ni u koju drugu svrhu niti se smiju na neki drugi način prenositi, otkrivati ili reproducirati bez prethodnog pisanog odobrenja tvrtke Illumina. Illumina ovim dokumentom ne prenosi nikakve licence zaštićene svojim pravom na patent, žig, autorskim pravom ili običajnim pravom ni slična prava bilo koje treće strane.

Kvalificirano i odgovarajuće obučeno osoblje mora se strogo i bez iznimki pridržavati uputa u ovom dokumentu da bi se zajamčila pravilna i sigurna upotreba proizvoda opisanih u njemu. Prije upotrebe proizvoda nužno je s razumijevanjem pročitati cjelokupan sadržaj dokumenta.

AKO UPUTE U DOKUMENTU NE PROČITATE U CIJELOSTI TE IH SE NE PRIDRŽAVATE BEZ IZNIMKI, MOŽE DOĆI DO OŠTEĆENJA PROIZVODA, OZLJEDA KORISNIKA ILI DRUGIH OSOBA I DO OŠTEĆENJA DRUGE IMOVINE TE SE TIME PONIŠTAVAJU SVA JAMSTVA ZA PROIZVODE.

ILLUMINA NE PREUZIMA ODGOVORNOST ZA ŠTETE NASTALE USLIJED NEPRAVILNE UPOTREBE PROIZVODA KOJI SU OPISANI U OVOM DOKUMENTU (UKLJUČUJUĆI DIJELOVE TIH PROIZVODA I SOFTVER).

© 2023 Illumina, Inc. Sva prava pridržana.

Svi su žigovi vlasništvo tvrtke Illumina, Inc. ili svojih vlasnika. Konkretno informacije o žigovima potražite na adresi www.illumina.com/company/legal.html.

Povijest revizija

Dokument	Datum	Opis promjene
200025238 v00	Veljača 2023	Početno izdanje.

Sadržaj

Povijest revizija	iii
Pregled	1
Metode analize	1
Izrada planirane obrade	5
Postavke	7
Datoteka manifesta	8
Filtriranje šuma (dodatna oprema)	9
Rezultati analize	9
FASTQ datoteke	10
BAM datoteke	11
VCF datoteke	11
Requeue Analysis (Ponovno stavi analizu u red čekanja)	17
Tehnička pomoć	18

Pregled

DRAGEN za Illumina DNK Prep with Enrichment Dx Aplikacija (DRAGEN za IDPE Dx) upotrebljava se za planiranje i izvođenje sekundarne analize biblioteka IDPE Dx generiranih za sekvenciranje na instrumentu NextSeq 550Dx .

DRAGEN za IDPE Dx podržava sekvenciranje do analize kada se upotrebljava s Illumina DNA Prep with Enrichment Dx Library PrepNextSeq 550Dx, i Illumina DRAGEN Server za NextSeq 550Dx.

Metode analize

DRAGEN za IDPE Dx provodi demultipleksiranje, generiranje FASTQ-a, mapiranje očitavanja i usklađivanje s referentnim genomom te očitavanje malih varijanti, ovisno o odabranom tijeku rada:

- Generiranje FASTQ
- Generiranje Germline FASTQ i VCF
- Generiranje Somatic FASTQ i VCF

NAPOMENA Kompresija ORA dostupna je za upotrebu sa sva tri tijeka rada. DRAGEN ORA Compression je softver za kompresiju bez gubitaka koji stvara datoteku s ekstenzijom Original Read Archive (*.ora). Format ora referentni je format kompresije za datoteke FASTQ i dizajniran je za vrlo brzu kompresiju/dekompresiju i visoki omjer kompresije.

Generiranje FASTQ

Sastavljeni slijedovi spremaju se u FASTQ datoteke po uzorku. FASTQ datoteke tekstualne su datoteke koje sadrže podatke o sekvenciranju i rezultate kvalitete samo za jedan uzorak. Za svaki uzorak generiraju se zasebne FASTQ datoteke po stazi protočne članke, po očitavanju sekvenciranja. Naziv uzorka naveden tijekom postavljanja obrade uključen je u naziv FASTQ datoteke. Datoteke FASTQ primarne su ulazne datoteke za usklađivanje. Prvi korak FASTQ generiranja je demultipleksiranje. Demultipleksiranjem se dodjeljuju klasteri koji prolaze filter uzorku usporedbom svake sekvence očitavanja indeksa sa sekvencama indeksa navedenim za obradu. U tom se koraku ne gledaju vrijednosti kvalitete. Očitavanja indeksa prepoznaju se u sljedećim koracima:

- Uzorci se numeriraju počevši od 1 na temelju redoslijeda kojim su navedeni za obradu.
- Broj uzorka 0 rezerviran je za klasterne koji nisu dodijeljeni uzorku.
- Klasteri se dodjeljuju uzorku kad se sekvenca indeksa točno podudara ili kad se nađe maksimalno jedno nepodudaranje po očitavanju indeksa.

Softver uključuje kompresiju ORA za komprimiranje FASTQ datoteka. Taj se format može dodatno omogućiti. Kada se koristi format ORA (*.ora), md5 checksum datoteka FASTQ sadržaja čuva se nakon ciklusa kompresije i dekompresije kako bi se osigurala kompresija bez gubitka podataka.

Mapiranje DNK i poravnanje

Nakon generacije FASTQ očitavanja se mapiraju i poravnavaju s referentnim genomom. Prva faza mapiranja je generiranje sjemena iz očitavanja, a zatim traženje točnih podudaranja u referentnom genomu. Ti se rezultati zatim rafiniraju pokretanjem cijelog Smith-Watermanova algoritama poravnanja na mjestima s najvećom gustoćom podudaranja sjemena. Ovaj dobro dokumentirani algoritam radi usporedbom svakog položaja očitavanja sa svim položajima kandidata u referentnom genomu. Te usporedbe odgovaraju matrici potencijalnih poravnanja između očitavanja i referentnog genoma. Za svaki od ovih položaja poravnanja kandidata, Smith-Watermanov algoritam generira rezultate koji se koriste kako bi se procijenilo je li najbolje poravnanje koje prolazi kroz tu ćeliju matrice nju doseglo podudaranjem ili neslaganjem nukleotida (dijagonalno pomicanje), brisanjem (vodoravno pomicanje) ili umetanjem (okomito kretanje). Podudaranje između očitavanja i reference daje bonus na rezultat, a nepodudaranje ili indel nameće kaznu. Ukupni najviši put ocjenjivanja kroz matricu je odabrano poravnanje. Algoritam je hardverski ubrzan na DRAGEN terenski programabilnim gate array karticama (FPGA). Referentni genom korišten u aplikaciji izrađen je iz UCSC hg19 FASTA s opcijom DRAGEN za kreiranje tablice ljestve na temelju alt-awareea.

Očitavanje DRAGEN Germline Variant

Očitavanje malih varijanti DRAGEN Germline mapira i poravnava DNK očitavanja kao ulazna i očitava polimorfizme pojedinačnih nukleotida (SNP) i umetanje ili deleciju (indeli) kroz kombinaciju detekcije u koloni i lokalnog *novog* skupa haplotipova. Da biste omogućili Očitavanje malih varijanti DRAGEN Germline, odaberite tijekom rada varijante zametnih stanica.

Pozivanje varijanti zametne linije se obično koristi za uzorke zametne linije za koje se zna da su ploidijske dvije. Referentna područja koja se mogu očitati najprije se identificiraju uz dovoljnu pokrivenost poravnanja. Unutar tih referentnih regija, brzo skeniranje sortiranih očitavanja prepoznaje aktivna područja koja su centrirana na stupcima nakupljanja s dokazima o varijanti. Aktivne regije podstavljene su s dovoljno konteksta za pokrivanje značajnog sadržaja nereferentnog sadržaja u blizini. Ako postoje dokazi o indelima, aktivne regije dobivaju dodatnu podstavu.

Usklađena očitavanja odrezuju se unutar svake aktivne regije i sastavljaju u De Bruijnov grafikon. Rubovi odrezanih očitavanja ponderirani su brojevima promatranja, s referentnom sekvencom kao okosnicom. Nakon čišćenja grafikona i pojednostavljenja, svi putovi pada ili porasta na grafikonu se izdvajaju kao haplotipovi kandidata. Svaki haplotip je Smith-Waterman poravnat s referentnim genomom kako bi se identificirale varijante koje predstavlja. Ovaj skup događaja može se pojačati otkrivanjem na temelju položaja. Za svaki par očitanih haplotipa, vjerojatnost $P(r|H)$ promatranja očitavanja, pod pretpostavkom da je haplotip pravi početni uzorak, procjenjuje se pomoću skrivenih Markovljevih modela (HMM).

Skeniranje po referentnom položaju u odnosu na aktivno područje, genotipovi kandidata formiraju se iz diploidnih kombinacija varijantnih događaja (SNP-ovi ili indeli). Za svaki događaj (uključujući referentni), uvjetna vjerojatnost $P(r|e)$ promatranja svakog preklapanja očitavanja procjenjuje se kao maksimum $P(r|H)$ za haplotipove koji podupiru događaj. Oni se kombiniraju u uvjetnu vjerojatnost $P(r|e1e2)$ za genotip (par događaja) i množe kako bi se dobila uvjetna vjerojatnost $P(R|e1e2)$ promatranja cijelog očitavanja nakupljanja. Pomoću Bayesove formule izračunava se posteriorna vjerojatnost $P(e1e2|R)$ svakog diploidnog genotipa i očitava se pobjednik.

DRAGEN za IDPE Dx primjenjuje automatsko filtriranje. Dodatne informacije potražite u [Napomene o VCF datoteci Germline Workflow na stranici 13](#).

Očitavanje DRAGEN Somatic Variant

Očitavanje malih varijanti DRAGEN Somatic mapira i poravnava DNK očitavanja kao ulazna i prepoznaje SNV-ove i indele putem lokalnog *novog* skupa haplotipova u aktivnom području. Da biste omogućili Očitavanje malih varijanti DRAGEN Somatic odaberite aplikaciju somatske varijante.

Otkrivanje somatskih varijanti obično se koristi za uzorke tumora. S ovim radnim procesom, DRAGEN ne stvara nikakve ploidne pretpostavke, što omogućuje otkrivanje niskofrekventnih alela. Za lokuse s pokrivenošću do 100x u uzorku tumora, DRAGEN ima prag otkrivanja pri učestalosti varijanti alela od 5%. Granica se skalira s povećanjem dubine na bazi po lokusu i prepolovi svaki put kada se pokrivenost udvostruči preko 100x. Referentna područja koja se mogu očitati najprije se identificiraju uz dovoljnu pokrivenost poravnanja. Unutar tih referentnih regija, skeniranje sortiranih očitavanja prepoznaje aktivne regije koje su centrirane na stupcima nakupljanja s dokazima o varijanti u očitanjima tumora. Aktivne regije podstavljene su s dovoljno konteksta za pokrivanje značajnog sadržaja nereferentnog sadržaja u blizini. Ako postoje dokazi o indelima, aktivne regije dobivaju dodatnu podstavu.

Usklađena očitavanja odrezuju se unutar svake aktivne regije i sastavljaju u De Bruijnov grafikon. Rubovi odrezanih očitavanja ponderirani su brojevima promatranja, s referentnom sekvencom kao okosnicom. Nakon čišćenja grafikona i pojednostavljenja, svi putovi pada ili porasta na grafikonu se izdvajaju kao haplotipovi kandidata. Svaki haplotip je Smith-Waterman poravnat s referentnim genomom kako bi se identificirale varijante koje predstavlja. Za svaki par očitavanja haplotipa, vjerojatnost $P(r|H)$ promatranja očitavanja, procjenjuje se pomoću skrivenih Markovljevih modela (HMM) pod pretpostavkom da je haplotip pravi početni uzorak.

Da bi se odredilo rezultat granice otkrivanja tumora (TLOD), Očitavanje malih varijanti DRAGEN Somatic prvo snima prema referentnom položaju za svaki somatski događaj kandidata kao i referentni događaj u aktivnom području. Uvjetna vjerojatnost $P(r|e)$ promatranja svakog preklapanja očitavanja procjenjuje se kao maksimum $P(r|H)$ za haplotipove koji podupiru događaj. Oni se kombiniraju u uvjetnu vjerojatnost $P(r|E)$ za hipotezu događaja, E, uključujući mješavinu referentnog i somatskog alela kandidata u rasponu mogućih frekvencija alela i množe se kako bi se dobila uvjetna vjerojatnost $P(R|E)$ promatranja cijelog očitavanja nakupljanja. Od tamo se izračunava TLOD rezultat kao dokaz da je alel ALT prisutan u uzorku tumora na određenom lokusu.

DRAGEN za IDPE Dx primjenjuje automatsko filtriranje. Dodatne informacije potražite u [Napomene o Somatic Workflow VCF datoteci](#) na stranici 15.

Izrada planirane obrade

Pomoću sljedećih koraka postavite obradu u Illumina Run Manager ili na NextSeq 550Dx ili pomoću preglednika na umreženom računalu. Ako želite uvesti podatke o uzorku, upotrijebite preglednik na umreženom računalu. Pogledajte Illumina Run Manager za Vodič za softver NextSeq 550Dx (broj dokumenta 200025239) za upute o pristupu Illumina Run Manager s umreženog računala.

Postoje dva različita načina za stvaranje novog planiranog ciklusa:

- **Import Run** (Uvoz obrade) – upotrijebite list s uzorcima iz postojeće obrade kao predložak za novu obradu. Pogledajte Illumina Run Manager za Vodič za softver NextSeq 550Dx (broj dokumenta 200025239) za informacije o tome kako uvesti obradu.
- **Create Run** (Izrada obrade) – ručno unesite parametre obrade. Upotrijebite sljedeće upute za izradu obrade.

NAPOMENA Potrebna polja za unos u korisničkom sučelju označena su zvjezdicom (*).

Aplikacija

1. Na kartici Planned (Planirano) zaslona Runs (Obrade) odaberite **Create Run** (Izradi obradu).
2. Odaberite aplikaciju DRAGEN za Illumina DNK Prep s Enrichment Dx, a zatim odaberite **Next** (Dalje).

Postavke obrade

1. Na zaslonu Run Settings (Postavke obrade) unesite naziv obrade. Naziv obrade prepoznaje obradu od sekvenciranja do analize.
2. **[Opcionalno]** Unesite opis obrade da biste je dodatno prepoznali.
3. Odaberite komplet(e) prilagodnika indeksa koji se koriste tijekom pripreme biblioteke.
4. Pregledajte duljinu očitavanja i po potrebi izmijenite. Očitavanje 1 i očitavanje 2 imaju zadanu vrijednost od 151 obrade. Indeks 1 i indeks 2 imaju fiksnu vrijednost od 10 obrada.
5. **[Opcionalno]** Unesite ID epruvete biblioteke.
6. Odaberite **Next** (Dalje).

Podaci o uzorku

Podaci o uzorku uključuju ID uzorka, položaj jažice (jasan položaj ploče za indeksiranje) i naziv biblioteke. Kada se koristi indeks A&B, položaj jažice također uključuje identifikator pločice.

Postoje dva načina za unos podataka o uzorku:

- **Import Samples** (Uvoz uzoraka) – upotrijebite datoteku predložka dostupnu za preuzimanje na zaslonu Sample Data (Podaci o uzorku).

- **Manually** (Ručno) — unesite podatke o uzorku izravno u tablicu na zaslonu Sample Data (Podaci o uzorku).

Uvoz uzoraka

Predložak (*.csv) dostupan je za preuzimanje na zaslonu Sample Data (Podaci o uzorcima) prilikom planiranja obrade u pomoću preglednika na umreženom računalu. Datoteka predložka nije dostupna za preuzimanje prilikom pristupanja Illumina Run Manager putem softvera NextSeq 550Dx operacijskog sustava. Da biste unijeli podatke o uzorku pomoću značajke Import Samples (Uvoz uzoraka), učinite sljedeće.

NAPOMENA Dovršite korake Run Settings (Postavke obrade) prije nego što nastavite.

1. Odaberite **Download Template** (Preuzmi predložak) za preuzimanje prazne CSV datoteke.
2. Iz datoteke predložka unesite podatke o uzorku, a zatim spremite datoteku. Naziv biblioteke nije obvezan.

NAPOMENA Prilikom korištenja indeksa A&B, podaci za stupac B moraju uključivati ploču i položaj jažice (indeksirani položaj ploče). Primjer: A-A01, A-A02, A-A03.

3. Odaberite **Import Samples** (Uvezi uzorke) i pogledajte datoteku predložka koja sadrži podatke o uzorcima iz prethodnog koraka.
4. Odaberite **Open** (Otvori), **Proceed** (Nastaviti), a zatim **Next** (Sljedeće).

NAPOMENA Promjena ID-a uzorka prije odabira Next (Sljedeće) može dovesti do pogreške. Završite postavljanje obrade prije mijenjanja kako biste izbjegli pogreške.

Ručni unos uzoraka

Pomoću tablice na zaslonu Sample Data (Podaci o uzorku) ručno unesite podatke o uzorku.

1. U polje Sample ID (ID uzorka) unesite jedinstveni ID uzorka.
2. Upotrijebite **Well Position** (Položaj jažice) (indeks A ili indeks B) ili **pločicu - položaj jažice** (indeks A&B) za odabir povezanog indeksa za uzorke. Polja i7 indeksa, indeksa 1, i5 indeksa i indeksa 2 automatski se popunjavaju.
3. **[Opcionalno]** Unesite naziv knjižnice.
4. Dodajte retke i po potrebi ponovite korake 1 – 3 dok svi uzorci ne budu dodani u tablicu. Možete dodati više redaka odjednom tako da najprije unesete broj redaka koji želite dodati, a zatim odaberete ikonu +. Redove možete ukloniti i odabirom okvira pored broja retka, a zatim klikom na ikonu smeća.
5. Odaberite **Next** (Dalje).

Postavke analize

1. Odaberite željeni tijek rada analize:
 - Generacija FASTQ
 - Generacija FASTQ-a i VCF-a za zametni tijek rada (potrebna je datoteka manifesta)
 - Generacija FASTQ-a i VCF-a za zametni tijek rada (potrebna je datoteka manifesta)
2. **[Opcionalno] Generirajte komprimirani FASTQ-ovi ORA** omogućeni su prema zadanim postavkama. Kompresija FASTQ ORA bez gubitka komprimira FASTQ datoteke do 5x u usporedbi s fastq.gz. Ako su poželjni nekomprimirani podaci (fastq.gz) odznačite **Generate ORA compressed FASTQs** (Generiraj komprimirane FASTQ-ove ORA).
3. Za zametne i somatske radne procese potrebna je datoteka manifesta. Na padajućem izborniku **Manifest File Selection** (Odabir datoteke manifesta) odaberite datoteku manifesta. Manifest je BED (*.bed) datoteka ograničena tablicom koja navodi nazive i lokacije ciljanih referentnih regija. Dodatne informacije potražite u odjeljku [Datoteka manifesta na stranici 8](#).
4. **[Opcionalno]** Za somatske radne procese koristite padajući izbornik **Noise File Selection** (Odabir datoteke šuma) za odabir datoteke sustavnog šuma. Može se odrediti datoteka BED (*.bed.gz) s razinom šuma specifičnom za centar za filtriranje sustavne šuma. Dodatne informacije potražite u odjeljku [Filtriranje šuma \(dodatna oprema\) na stranici 9](#).
5. Odaberite **Next** (Dalje).

Obrada Pregled

1. Na zaslону Review (Pregled) pregledajte informacije unesene na zaslonima Run Settings (Postavke obrade), Sample Data (Podaci o uzorcima) i Analysis Settings (Postavke analize).
2. Odaberite **Save** (Spremi).
Obrada se sprema na kartici Planned (Planirano) na zaslonu Runs (Obrade).

Postavke

Za prikaz ili promjenu postavki aplikacije DRAGEN za IDPE Dx najprije odaberite ikonu Applications (Aplikacije) na glavnom zaslonu. Zatim odaberite aplikaciju koju želite pregledati ili promijeniti. Za promjenu postavki potreban je administratorski račun.

Konfiguracija

Zaslon konfiguracije prikazuje sljedeće postavke aplikacije:

- **Library Prep Kits** (Kompleti za pripremu knjižnice)— Prikazuje zadani komplet za pripremu knjižnice za aplikaciju. Ova se postavka ne može promijeniti.
- **Index Adapter Kits** (Kompleti prilagodnika indeksiranja)— Prikazuje zadani komplet prilagodnika indeksiranja za aplikaciju. Ova se postavka ne može promijeniti.

- **Read Lengths** (Dužine očitavanja)— Duljine očitavanja za aplikaciju postavljene su prema zadanim postavkama na 151, ali se mogu promijeniti tijekom izrade ciklusa obrade.
- **Manifest and Noise Files** (Datoteke manifesta i šuma) – prenesite i promijenite postavke za datoteke manifesta i šuma.
 - Odaberite **Upload File** (Prijenos datoteke) za prijenos datoteka za korištenje u analizi.
 - Odaberite radijski gumb **Default** (Zadano) da biste postavili datoteku kao zadanu datoteku manifesta ili šuma koja je odabrana tijekom stvaranja obrade kada je odabrana aplikacija.
 - Odaberite potvrdni okvir **Enabled** (Omogućeno) da biste na padajućem izborniku postavili datoteku za prikaz tijekom izrade ciklusa obrade.

Dozvole

Upotrijebite potvrdne okvire na zaslону Permissions (Dozvole) da biste upravljali pristupom korisnika aplikaciji.

Datoteka manifesta

Kada koristite DRAGEN za IDPE Dx, potreban je ulaz datoteke manifesta za sljedeće radne procese:

- Germline FASTQ i VCF generiranje za tijek rada sa spolnim stanicama
- Somatic FASTQ i VCF generiranje za somatski tijek rada

Manifest je BED (*.bed) datoteka s podacima odvojenim tabulatorom koja navodi nazive i lokacije ciljanih referentnih regija. Glavni dio datoteke manifesta je odjeljak Regions (Regije) i treba sadržavati sljedeće stupce podataka:

Stupac	Opis
Naziv	Jedinstveni korisnički određeni naziv za cilj
Kromosom	Mjesto kromosoma (npr. chr10, chr5 itd.)
Pokreni	1-temeljni indeks za početni položaj cilja
Zaustavljanje	1-temeljni indeks za zaustavni položaj cilja
Duljina uzvodne sonde	Duljina uzvodne sonde. Za DRAGEN za IDPE Dx aplikaciju, to treba postaviti na 0.
Duljina nizvodne sonde	Duljina nizvodne sonde. Za DRAGEN za IDPE Dx aplikaciju, to treba postaviti na 0.

NAPOMENA Za analizu je potreban valjan format datoteke manifesta. DRAGEN će zaustaviti analizu ako datoteka manifesta nije valjana.

Filtriranje šuma (dodatna oprema)

Sustavni filter šuma dostupan je za prepoznavanje somatskih varijanti i može se koristiti za smanjenje lažno pozitivnih poziva uzimajući u obzir buku specifičnu za lokaciju. Sustavna datoteka šuma generira se prvo prikupljanjem približno 50 normalnih uzoraka (po mogućnosti specifičnih za panel, pripremu biblioteke i sekvencu) i zatim zbrojem učestalosti alela ispod 30 % na svakom mjestu s dovoljnom pokrivenošću dijeli se s ukupnim brojem uzoraka (za frekvencije alela iznad 30 % se pretpostavlja da su varijante zametnih stanica, a ne buka). Nakon generiranja vrijednosti šuma, somatske varijante otkrivene na tom mjestu filtrirat će se.

Filter se može koristiti u Tumor-Normal (Tumorsko-normalnom) načinu rada, ali je posebno koristan samo za obrade tumora gdje podudarna normalna vrijednost nije dostupna. Sustavna datoteka mora koristiti BED datoteku koja ima nastavak datoteke (`*.bed.gz`) i mora sadržavati četiri stupca: Kromosom, pokretanje, kraj i razine šuma specifične za lokaciju za svaki red. Sustavno filtriranje šuma nije obavezno.

Rezultati analize

Obrade koje su trenutačno u tijeku prikazuju se na kartici Active (Aktivno). Dovršeni ciklusi prikazuju se na kartici Completed (Dovršeno). DRAGEN za IDPE Dx stvara jedinstvenu mapu analize za svaku analizu koja je odvojena od mape koja sadrži podatke sekvenciranja. Mapa analize uključuje sljedeće informacije:

- Korištena datoteka manifesta
- Verzija softvera
- ID-ovi uzoraka
- Ukupne usklađene baze
- Postotak usklađenih očitavanja po uzorku
- Broj SNV-ova očitanih po uzorku
- Broj indela očitanih po uzorku
- Statistika pokrivenosti

Izlazne datoteke analize

Lokacija mape analize određena je postavkom External Storage for Analysis Results (Vanjska pohrana za rezultate analize). Dodatne informacije o postavci External Storage for Analysis Results (Vanjska pohrana za rezultate analize) potražite u dokumentu Illumina Run Manager za Vodič za softver NextSeq 550Dx (broj dokumenta 200025239).

Na zaslону Run Details (Pojednosti obrade) polje External Location (Vanjska lokacija) pruža put za podatke sekvenciranja. Jedinstveni naziv mape analize nalazi se u polju Analysis Output Folder (Mapa izlaza analize) na zaslону Run Details (Pojednosti obrade). Koje će se točno datoteke generirati ovisno o tome koji se tijekom rada analize koristi. Aplikacija generira sljedeće izlazne datoteke za analizu.

NAPOMENA Ako se prilikom pristupanja izlaznim datotekama za analizu pojavi pogreška ograničenja maksimalne duljine puta datoteke, pokušajte premjestiti datoteku na lokaciju kraćeg puta ili upotrijebite drugi način za otvaranje datoteke.

Izlazna datoteka	Opis
Sažetak izvješća varijante (* .pdf)	Sadrži sažetak informacija o datoteci, verzije softvera, informacije o uzorku, statistiku razine čitanja i sažetke SNV-a, umetanja, brisanja i pokrivenosti. Samo spolni i somatski tijekovi rada generiraju izvješća o varijanti.
FASTQ (* .fastq.gz ili * .fastq.ora)	Sporadne datoteke koje sadrže očitavanja baza za koja postoji ocjena kvalitete. FASTQ datoteke primarne su ulazne datoteke u koraku usklađivanja. Kada se odabere kompresija ORA, koristi se nastavak datoteke * .fastq.ora.
Usklađivanje BAM datoteka (* .bam)	Sadrže usklađena očitavanja za određeni uzorak.
Genomske VCF datoteke (* .gvcf.gz)	Sadrži genotip za svaki položaj, očitani kao varijanta ili kao referenca.
VCF datoteke (* .vcf.gz)	Sadrže varijante očitane na svakom položaju.
Pokrenite metričko izvješće (* .csv)	Sadrži mjerne podatke kvalitete obrade, uključujući neindeksiranu ukupnu iskoristivost i ocjenu kvalitete za Q30.

FASTQ datoteke

FASTQ (* .fastq.gz, * .fastq.ora) je format datoteke utemeljen na tekstu koji sadrži očitavanja baza i vrijednosti kvalitete po očitavanju. Svaka datoteka sadrži sljedeće informacije:

- Identifikator uzorka
- Sekvence
- Phred bodovanja kvalitete u formatu kodiranom prema standardu ASCII + 33

Identifikator uzorka formatiran je na sljedeći način:

```
@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 datoteke

BAM datoteka (*.bam) je komprimirana binarna verzija SAM (mapa poravnanja sekvenci) datoteke koja se upotrebljava za predstavljanje usklađenih sekvenci veličine do 128 Mb. BAM datoteke upotrebljavaju oblik imenovanja datoteka `SampleName_S#.bam`, pri čemu je # broj uzorka određen redoslijedom kojim su uzorci navedeni za obradu. U multinodnom načinu rada, S# je postavljen na S1, bez obzira na redoslijed uzorka.

BAM datoteke sadrže odjeljak zaglavlja i odjeljak usklađivanja:

- **Header** (Zaglavlje) – sadrži informacije o cijeloj datoteci kao što su naziv uzorka, dužina uzorka i metoda usklađivanja. Usklađivanja u odjeljku usklađivanja povezana su s određenim informacijama u odjeljku zaglavlja.
- **Alignments** (Usklađivanja) – sadrži naziv očitavanja, sekvencu očitavanja, kvalitetu očitavanja, informacije o usklađivanju i prilagođene oznake. Naziv očitavanja obuhvaća kromosom, početnu koordinatu, kvalitetu usklađivanja i podudarni niz opisnika.

Sekcija usklađivanja obuhvaća sljedeće informacije za svako očitavanje ili par očitavanja:

- AS: kvaliteta usklađivanja s uparenim krajevima.
- RG: grupa očitavanja koja upućuje na broj očitavanja za određeni uzorak.
- BC: oznaka crtičnog koda koja upućuje na ID demultipleksiranog uzorka povezanog s očitanjem.
- SM: kvaliteta usklađivanja s jednim krajem.
- XC: podudarni niz opisnika.
- XN: oznaka naziva amplikona koja bilježi ID amplikona povezanog s očitanjem

BAM datoteke indeksa (*.bam.bai) sadrže indeks odgovarajuće BAM datoteke.

VCF datoteke

Format očitavanja varijante (*.vcf) sadrži informacije o varijantama pronađenim na određenim mjestima u referentnom genomu.

Zaglavlje VCF datoteke sastoji se od verzije formata datoteke VCF, verzije očitavanja varijante te navodi oznake korištene u ostatku datoteke. VCF zaglavlje obuhvaća i datoteku referentnog genoma te BAM datoteku. Zadnji redak zaglavlja sadrži naslove stupaca za retke podataka. Svaki redak podataka u VCF datoteci sadrži informacije o jednoj varijanti.

Tablica 1 Zaglavlja VCF datoteka

Zaglavlje	Opis
CHROM	Kromosom referentnog genoma. Kromosomi se pojavljuju istim redoslijedom kao u referentnoj datoteci FASTA.
POS	Jednobazni položaj varijante u referentnom kromosomu. Za jednonukleotidne varijante (SNV-ove), ovaj je položaj referentna baza s varijantom. Za indele, ovaj je položaj referentna baza koja je neposredno ispred varijante.
ID	Rs (referentni SNP) broj varijante dobiven iz datoteke <code>dbSNP.txt</code> , ako je to primjenljivo. Ako na tom mjestu postoji više rs brojeva, popis je odijeljen točkama-zarezima. Ako na tom mjestu ne postoji dbSNP unos, primjenjuje se oznaka ('.') za vrijednost koja nedostaje.
REF	Referentni genotip. Primjerice, brisanje jednog T predstavljeno je kao referentni TT i alternativni T. Jedna varijanta nukleotida A na T predstavljena je kao referentni A i alternativni T.
ALT	Aleli koji se razlikuju od referentnog očitavanja. Primjerice, umetanje jednog T predstavljeno je kao referentni A i alternativni AT. Jedna varijanta nukleotida A na T predstavljena je kao referentni A i alternativni T.
QUAL	Phred bodovanje kvalitete koje dodjeljuje očitavatelj varijante. Veći broj bodova upućuje na veću pouzdanost varijante i manju mogućnost pogrešaka. Za bodovanje kvalitete Q procijenjena vjerojatnost pogreške iznosi $10^{-(Q/10)}$. Primjerice, skup od Q30 očitavanja ima omjer pogrešaka od 0,1 %. Mnogi očitavatelji varijanti boduju kvalitetu na temelju svojih statističkih modela, koji visoko koreliraju s uočenim omjerom pogrešaka.

Tablica 2 Napomene o VCF datoteci Germline Workflow

Zaglavlje	Opis
FILTAR	<p>Ako su svi filtri prošli, u stupcu filtra piše PASS (PROLAZ). Mogući unosi FILTERA uključuju:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL—Primijenjuje se ako varijanta SNP QUAL rezultata ne zadovoljava granične vrijednosti • DRAGENIndelHardQUAL—Primijenjuje se ako indel varijanta QUAL rezultata ne zadovoljava granične vrijednosti • LowDepth—Lokacija je filtrirana jer dubina pokrivenosti ne zadovoljava granične vrijednosti • LowGQ—Centar filtriran jer kvaliteta genotipa ne zadovoljava prag • PloidyConflict—Očitavanje genotipa od pozivatelja varijanti nije u skladu s ploidijom kromosoma • base_quality—Lokacija je filtrirana jer srednja osnovna kvaliteta alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti • filtered_reads—Lokacija je filtrirana jer je prevelik dio očitavanja filtriran • fragment_length—Lokacija je filtrirana jer apsolutna razlika između srednje duljine fragmenta alt očitavanja i srednje duljine fragmenta ref očitavanja na tom lokusu prelazi granične vrijednosti • low_deep—Lokacija je filtrirana jer je dubina očitavanja preniska • low_frac_info_reads—Lokacija je filtrirana jer je udio informativnih očitavanja ispod granične vrijednosti • low_normalna_depth—Lokacija je filtrirana jer je normalna dubina očitavanja uzorka preniska • long_indel—Lokacija je filtrirana jer je duljina indela preduga • mapping_quality—Lokacija je filtrirana jer srednja kvaliteta mapiranja alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti • multiallelic—Lokacija je filtrirana jer više od dva alt alela prolaze tumorski LOD • non_homref_normal—Lokacija je filtrirana jer genotip normalnog uzorka nije homozigotna referenca • no_reliable_supporting_read—Lokacija je filtrirana jer ne postoji pouzdano prateće somatsko očitavanje • panel_of_normals—primijećeno u najmanje jednom uzorku u panelu normalne vcf datoteke • read_position—Lokacija je filtrirana jer je medijan udaljenosti između početka/kraja očitavanja i ovog lokusa ispod granične vrijednosti • RMxNRepeatRegion—Lokacija je filtrirana jer je cijeli ili dio alelnih varijanti ponavljen u referenci • strand_artifact—Lokacija je filtrirana zbog jake pristranosti niza • str_contraction—Lokacija je filtrirana zbog sumnje na pogrešku PCR-a gdje je alt alel jedna ponovljena jedinica manja od referentne • too_few_supporting_reads—Lokacija je filtrirana jer ima premalo popratnih očitavanja u uzorku tumora • weak_evidence—Rezultat somatske varijante ne zadovoljava granične vrijednosti

Zaglavlje	Opis
INFO	<p>Mogući INFO unosi uključuju:</p> <ul style="list-style-type: none"> • AC—Broj alela u genotipima za svaki ALT alel, istim redoslijedom kojim su navedeni. • AF—Frekvencija alela za svaki ALT alel, istim redoslijedom kojim su navedeni. • AN—Ukupan broj alela u očitanim genotipima. • DB—dbSNP članstvo. • FS—Phred bodovanje p-vrijednosti pomoću Fisherovog egzaktnog testa za otkrivanje pristranosti niza. • QD—Pouzdanost varijante / kvaliteta prema dubini. • R2_5P_bias—Ocjena na temelju sklonosti niza i udaljenosti od 5 glavnih krajeva. • SOR—Simetrični omjer vjerojatnosti 2x2 tablice nepredviđenosti za otkrivanje pristranosti niza. • DP—Približna dubina očitavanja (informativna i neinformativna); neka očitavanja možda su filtrirana na temelju mapq rezultata itd. • END—Zaustavni položaj intervala. • FractionInformativeReads—Dio informativnih očitavanja od ukupnog broja očitavanja. • MQ—RMS kvaliteta mapiranja. • MQRankSum—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref očitavanje kvalitete mapiranja. • ReadPosRankSum—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref pristranost očitavog položaja. • SOMATIC—Najmanje jedna varijanta na tom položaju je somatska.
FORMAT	<p>U stupcu Format navedena su polja, odvojena dvotočkama. Npr. GT:GQ. Dostupna su sljedeća polja:</p> <ul style="list-style-type: none"> • AD—Dubine alela (koje sadrže samo informativne očitavanja od ukupnog broja očitavanja) za ref i alt alele navedenim redoslijedom. • AF—Frakcije alela za alt alele navedenim redoslijedom. • DP—Približna dubina očitavanja (filtriraju se očitavanja s MQ=255 ili lošim parovima). • F1R2—Broj očitavanja u parnom usmjerenju F1R2 koja podržavaju svaki alel. • F2R1—Broj očitavanja u parnom usmjerenju F2R1 koja podržavaju svaki alel. • GT—Genotip. 0 odgovara referentnoj bazi, 1 odgovara prvom unosu u stupac ALT itd. Kosa crta (/) ukazuje da nema dostupnih informacija o fazama. • MB—Statistika komponente po uzorku za otkrivanje sklonosti niza. • PS—Informacije o ID-u fizičkog faziranja, pri čemu svaki jedinstveni ID unutar određenog uzorka (ali ne između uzoraka) povezuje zapise unutar grupe za faziranje. • SB—Statistika komponenti po uzorku koja se sastoji od Fisherovog egzaktnog testa za otkrivanje pristranosti niza. • SQ—Somatska kvaliteta.
UZORAK	Stupac uzorka sadrži vrijednosti navedene u stupcu FORMAT.

Tablica 3 Napomene o Somatic Workflow VCF datoteci

Zaglavlje	Opis
FILTAR	<p>Ako su svi filtri prošli, u stupcu filtra piše PASS (PROLAZ). Mogući unosi FILTERA uključuju:</p> <ul style="list-style-type: none"> • base_quality—Lokacija je filtrirana jer srednja osnovna kvaliteta alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti • filtered_reads—Lokacija je filtrirana jer je prevelik dio očitavanja filtriran • fragment_length—Lokacija je filtrirana jer apsolutna razlika između srednje duljine fragmenta alt očitavanja i srednje duljine fragmenta ref očitavanja na tom lokusu prelazi granične vrijednosti • low_deep—Lokacija je filtrirana jer je dubina očitavanja preniska • low_frac_info_reads—Lokacija je filtrirana jer je udio informativnih očitavanja ispod granične vrijednosti • low_normalna_depth—Lokacija je filtrirana jer je normalna dubina očitavanja uzorka preniska • long_indel—Lokacija je filtrirana jer je duljina indela preduga • mapping_quality—Lokacija je filtrirana jer srednja kvaliteta mapiranja alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti • multiallelic—Lokacija je filtrirana jer više od dva alt alela prolaze tumorski LOD • non_homref_normal—Lokacija je filtrirana jer genotip normalnog uzorka nije homozigotna referenca • no_reliable_supporting_read—Lokacija je filtrirana jer ne postoji pouzdano prateće somatsko očitavanje • panel_of_normals—primijećeno u najmanje jednom uzorku u panelu normalne vcf datoteke • read_position—Lokacija je filtrirana jer je medijan udaljenosti između početka/kraja očitavanja i ovog lokusa ispod granične vrijednosti • RMxNRepeatRegion—Lokacija je filtrirana jer je cijeli ili dio alelnih varijanti ponavljen u referenci • strand_artifact—Lokacija je filtrirana zbog jake pristranosti niza • str_contraction—Lokacija je filtrirana zbog sumnje na pogrešku PCR-a gdje je alt alel jedna ponovljena jedinica manja od referentne • too_few_supporting_reads—Lokacija je filtrirana jer ima premalo popratnih očitavanja u uzorku tumora • weak_evidence—Rezultat somatske varijante ne zadovoljava granične vrijednosti • sistematska_buka—Lokacija filtrirana na temelju dokaza sustavnog šuma u normalnim uvjetima

Zaglavlje	Opis
INFO	<p>Mogući INFO unosi uključuju:</p> <ul style="list-style-type: none"> • DP—Približna dubina očitavanja (informativna i neinformativna); neka očitavanja možda su filtrirana na temelju mapq rezultata itd. • END—Zaustavni položaj intervala. • FractionInformativeReads—Dio informativnih očitavanja od ukupnog broja očitavanja. • MQ—RMS kvaliteta mapiranja. • MQRankSum—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref očitavanje kvalitete mapiranja. • ReadPosRankSum—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref pristranost očitavog položaja. • AQ—Rezultat sustavnog šuma. • hotspot—Poznato somatsko mjesto, koristi se za povećanje pouzdanosti očitavanja. • SOMATIC—Najmanje jedna varijanta na tom položaju je somatska.
FORMAT	<p>U stupcu Format navedena su polja, odvojena dvotočkama. Npr. GT:GQ. Dostupna su sljedeća polja:</p> <ul style="list-style-type: none"> • AD—Dubine alela (koje sadrže samo informativne očitavanja od ukupnog broja očitavanja) za ref i alt alele navedenim redoslijedom. • AF—Frakcije alela za alt alele navedenim redoslijedom. • DP—Približna dubina očitavanja (filtriraju se očitavanja s MQ=255 ili lošim parovima). • F1R2—Broj očitavanja u parnom usmjerenju F1R2 koja podržavaju svaki alel. • F2R1—Broj očitavanja u parnom usmjerenju F2R1 koja podržavaju svaki alel. • GP—Posteriorne vjerojatnosti prema Phred bodovanju za genotipove prema definiciji u VCF specifikaciji. • GQ—Kvaliteta genotipa. • GT—Genotip. 0 odgovara referentnoj bazi, 1 odgovara prvom unosu u stupac ALT itd. Kosa crta (/) ukazuje da nema dostupnih informacija o fazama. • MB—Statistika komponente po uzorku za otkrivanje sklonosti niza. • PL—Normalizirano, Phred bodovanje vjerojatnosti za genotipove kako je definirano u VCF specifikaciji. • PRI—Prethodne vjerojatnosti prema Phred bodovanju za genotipove. • PS—Informacije o ID-u fizičkog faziranja, pri čemu svaki jedinstveni ID unutar određenog uzorka (ali ne između uzoraka) povezuje zapise unutar grupe za faziranje. • SB—Statistika komponenti po uzorku koja se sastoji od Fisherovog egzaktnog testa za otkrivanje pristranosti niza. • SQ—Somatska kvaliteta.
UZORAK	Stupac uzorka sadrži vrijednosti navedene u stupcu FORMAT.

Genomske VCF datoteke

Genomske VCF (*.gvcf.gz) datoteke slijede niz dogovorenih pravila za predstavljanje svih mjesta u genomu u relativno kompaktnom obliku. gVCF datoteke obuhvaćaju sva mjesta unutar područja od interesa u jednoj datoteci za svaki uzorak. gVCF datoteka pokazuje neočitavanja na položajima koji ne zadovoljavaju sve filtre. Oznaka genotipa (GT) ./. ukazuje na neočitavanje.

Requeue Analysis (Ponovno stavi analizu u red čekanja)

Ako analiza nije bila uspješna ili želite ponoviti analizu obrade s drugačijim postavkama, možete ponovno staviti analizu u red čekanja ako je bila zaustavljena. Da biste ponovno stavili analizu u red čekanja, učinite sljedeće:

1. Na zaslonu Run (Obrada) odaberite karticu Completed (Dovršeno), a zatim odaberite naziv obrade za ponovnu analizu.
Ako je prethodno provedeno Requeue Analysis (Ponovno stavi analizu u red čekanja), odaberite naziv obrade Parent Run (Prethodna obrada).
2. Na zaslonu Run Details (Pojedinsti obrade) nakon sekvenciranja odaberite **Requeue Analysis** (Analiza u red čekanja).
3. Odaberite opciju:
 - Ponovno stavi analizu u red čekanja bez promjena
 - Uredite postavke obrade i ponovno stavite analizu u red čekanja
 - Requeue Analysis (Ponovno stavi analizu u red čekanja) s drugom primjenom
4. Potvrdite da je lokacija na kojoj se trenutno nalaze podaci dobiveni sekvenciranjem navedena u polju **Sequencing data file path** (Put datoteke podataka sekvenciranjem).

NAPOMENA Put do podataka sekvenciranja trebao bi odgovarati putanji u postavci vanjske pohrane za analizu rezultata. Pogledajte Illumina Run Manager za Vodič za softver NextSeq 550Dx (broj dokumenta 200025239) za informacije o promjeni vanjskog puta pohrane.

5. Unesite razlog za ponovnu analizu.
6. Odaberite **Requeue Analysis** (Ponovno stavi analizu u red čekanja).
7. Uredite željene promjene Run Settings (Postavke obrade, Sample Data (Podaci o uzorcima) i Analysis Settings (Postavke analize).
8. Odaberite **Save** (Spremi). Analiza počinje uz korištenje trenutnih parametara analize.

Tehnička pomoć

Ako vam je potrebna tehnička pomoć, obratite se Illumina Tehnička podrška.

Web-mjesto: www.illumina.com

Adresa e-pošte: techsupport@illumina.com

Sigurnosno-tehnički listovi (SDS-ovi) – dostupni su na web-mjestu tvrtke Illumina na adresi support.illumina.com/sds.html.

Dokumentacija o proizvodima – dostupna je za preuzimanje na web-mjestu support.illumina.com.



Illumina

5200 Illumina Way

San Diego, California 92122 SAD

+1.800.809.ILMN (4566)

+1.858.202.4566 (izvan Sjeverne Amerike)

techsupport@illumina.com

www.illumina.com

CE

IVD

EC REP



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

Australski sponsor

Illumina Australia Pty Ltd

Nursing Association Building

Level 3, 535 Elizabeth Street

Melbourne, VIC 3000

Australija

ZA IN VITRO DIJAGNOSTIKU.

© 2023 Illumina, Inc. Sva prava pridržana.

illumina[®]