

DRAGEN for Illumina DNA Prep with Enrichment Dx on NextSeq 550Dx

Användarguide för applikation

TILLHÖR ILLUMINA

Dokumentnr 200025238 v00

Februari 2023

FÖR IN VITRO-DIAGNOSTISKT BRUK.

Dokumentet och dess innehåll tillhör Illumina, Inc. och dess dotterbolag ("Illumina") och är endast avsett för användning enligt avtal i samband med kundens bruk av produkterna som beskrivs häri. Allt annat bruk är förbjudet. Dokumentet och dess innehåll får ej användas eller distribueras i något annat syfte och/eller återges, delges eller reproduceras på något vis utan föregående skriftligt tillstånd från Illumina. I och med detta dokument överlåter Illumina inte någon licens som hör till dess patent, varumärke eller upphovsrätt, eller i enlighet med rättspraxis eller liknande tredjepartsrättigheter.

Instruktionerna i detta dokument ska följas noggrant och uttryckligen av kvalificerad och lämpligt utbildad personal för att säkerställa rätt och säker produktanvändning i enlighet med beskrivningen häri. Hela innehållet i dokumentet ska läsas och förstås i sin helhet innan produkten (produkterna) används.

UNDERLÅTENHET ATT LÄSA OCH FÖLJA ALLA INSTRUKTIONER HÄRI I SIN HELHET KAN MEDFÖRA SKADA PÅ PRODUKTEN/PRODUKTERNA, PERSONSKADA, INKLUSIVE SKADA PÅ ANVÄNDAREN/ANVÄNDARNA ELLER ANDRA PERSONER SAMT SKADA PÅ ANNAN EGENDOM, OCH LEDER TILL ATT EVENTUELL GARANTI FÖR PRODUKTEN/PRODUKTERNA BLIR OGILTIG.

ILLUMINA KAN INTE ÅLÄGGAS NÅGOT ANSVAR SOM UPPKOMMER GENOM FELAKTIG ANVÄNDNING AV PRODUKTERNA SOM BESKRIVS HÄRI (INKLUSIVE DELAR DÄRI ELLER PROGRAM).

© 2023 Illumina, Inc. Med ensamrätt.

Alla varumärken tillhör Illumina, Inc. eller respektive ägare. Specifik varumärkesinformation finns på www.illumina.com/company/legal.html.

Revisionshistorik

Dokument	Datum	Beskrivning av ändring
200025238 v00	Februari 2023	Första utgåvan.

Innehållsförteckning

Revisionshistorik	iii
Översikt	1
Analysmetoder	1
Skapa en planerad körning	5
Inställningar	7
Manifestfil	8
Brusfiltrering (valfritt)	9
Analysutdata	9
FASTQ-filer	10
BAM-filer	11
VCF-filer	11
Repetera analys	17
Teknisk hjälp	18

Översikt

DRAGEN for Illumina DNA Prep with Enrichment Dx Applikationen (DRAGEN for IDPE Dx) används för att planera och utföra sekundär analys av IDPE Dx-bibliotek, som genererats för sekvensering på NextSeq 550Dx.

DRAGEN for IDPE Dx stöder sekvensering för analys när den används med Illumina DNA Prep with Enrichment Dx bibliotekspreparering, NextSeq 550Dx och Illumina DRAGEN Server for NextSeq 550Dx.

Analysmetoder

DRAGEN for IDPE Dx utför demultiplexering, FASTQ-generering, läsmappning, inriktning till ett referensgenom och liten variantbestämning beroende på de valda arbetsflödena:

- FASTQ-generering
- Germline FASTQ- och VCF-generering
- Somatic FASTQ- och VCF-generering

OBS! ORA-kompression är tillgänglig för användning med alla tre arbetsflödena. DRAGEN ORA-kompression är ett helt förlustfritt komprimeringsprogram, som skapar en fil med ett Original Read Archive-tillägg (*.ora). Ora-formatet är ett referensbaserat komprimeringsformat för FASTQ-filer, och är utformat för mycket snabb komprimering/dekomprimering och högt komprimeringsförhållande.

FASTQ-generering

De sammansatta sekvenserna skrivs till FASTQ-filer per prov. FASTQ-filer är textfiler som innehåller sekvenseringsdata och kvalitetsresultat för endast ett prov. För varje prov genereras separata FASTQ-filer per flödescellfil, per sekvensläsning. Namnet på provet som specificerats under körningsinstallationen ingår i FASTQ-filnamnet. FASTQ-filer är inpassningens primära indata. Det första steget i FASTQ-genereringen är demultiplexering. Demultiplexering tilldelar kluster som skickar filter till ett prov genom att jämföra varje indexlässekvens med de indexsekvenser som anges för körningen. Inga kvalitetsvärden beaktas i det här steget. Indexavläsningar identifieras med följande steg:

- Proverna numreras i den ordning som de anges för körningen, med början på 1.
- Provnummer 0 är reserverat för kluster som inte tilldelats till något prov.
- Kluster tilldelas ett prov när indexsekvensen matchar exakt eller när det finns upp till en enda felmatchning per indexläsning.

Programvaran inkluderar ORA-komprimering för att komprimera FASTQ-filer. Detta format kan valfritt aktiveras. När du använder formatet ORA (*.ora) bevaras md5-kontrollsumman för FASTQ-innehållet efter en komprimerings- och dekomprimeringscykel för att säkerställa en förlustfri komprimering.

DNA-kartläggning och justering

Efter FASTQ-generering mappas avläsningarna och justeras till ett referensgenom. Det första steget i kartläggningen är att generera frön från avläsningen och sedan leta efter exakta matchningar i referensgenomet. Dessa resultat förfinas sedan genom att köra fullständiga Smith-Waterman-inriktningar på platserna med den högsta tätheten av utsäde. Denna väldokumenterade algoritm fungerar genom att jämföra varje position i avläsningen med alla kandidatpositioner i referensen. Dessa jämförelser motsvarar en matris av potentiella anpassningar mellan läsning och referens. För var och en av dessa kandidatpositioner genererar Smith-Waterman poäng som används för att utvärdera om den bästa inriktningen som passerar genom den matriscellen når den genom en nukleotidmatchning eller felmatchning (diagonal rörelse), en deletion (horisontell rörelse) eller en insertion (vertikal rörelse). En matchning mellan läsning och referens ger en bonus på poängen, och en felmatchning eller indel ger ett straff. Den övergripande vägen för högsta poäng genom matrisen är den valda inriktningen. Algoritmen är maskinvara, som accelereras på DRAGEN-kort med fältprogrammerbar grindmatris (FPGA). Referensgenomet, som används i appen, skapas från UCSC hg19 FASTA med DRAGEN-alternativet, för att skapa en överlyftsbasead alt-aware hash-tabell.

DRAGEN Bestämning av könszellvariant

DRAGEN Germline Small Variant Caller tar kartade och justerade DNA-avläsningar som indata och bestämmer genetiska enstaka nukleotidpolymorfismer (SNP) och insertion eller deletion (indels) genom en kombination av kolumnvis detektering och lokal *de novo*-sammansättning av haplotyper. För att aktivera DRAGEN Germline Small Variant Caller, välj arbetsflödet för könszellvariant.

Bestämning av könszellvariant används vanligtvis för könszellprover där ploidin är känd för att vara två. Bestämbara referensregioner identifieras först med tillräcklig inriktningstäckning. Inom dessa referensregioner identifierar en snabb skanning av de sorterade avläsningarna aktiva regioner, som är centrerade på pileup-kolumner med bevis på en variant. De aktiva regionerna är utfyllda med tillräckligt med sammanhang för att täcka betydande, icke-referensinnehåll i närheten. Om det finns bevis på indels får de aktiva regionerna ytterligare utfyllnad.

Justerade avläsningar klipps inom varje aktiv region och sätts samman till en De Bruijn-graf. Kanterna på de klippta avläsningarna viktas av observationsräkningar, med referenssekvensen som en ryggrad. Efter viss grafrensning och förenkling extraheras alla källa-till-sänkvägar som kandidathaplotyper. Varje haplotyp är Smith-Waterman anpassad till referensgenomet för att identifiera de varianter den representerar. Denna uppsättning händelser kan utökas med en positionsbaserad detektering. För varje läs-haplotyp uppskattas sannolikheten $P(r|H)$ för att observera läsningen, förutsatt att haplotypen är det sanna startprovet, med hjälp av en par-dold Markov-modell (HMM).

Genom att skanna genom referensposition över den aktiva regionen bildas kandidatgenotyper från diploida kombinationer av varianthändelser (SNP eller indels). För varje händelse (inklusive referens) uppskattas den villkorade sannolikheten $P(r|e)$ för att observera varje överlappande avläsning som det maximala $P(r|H)$ för haplotyper som stöder händelsen. Dessa kombineras till den villkorliga sannolikheten $P(r|e1e2)$ för en genotyp (händelsepar) och multipliceras för att ge den villkorade sannolikheten $P(R|e1e2)$ för att observera hela den lästa pileupen. Med hjälp av Bayes formel beräknas den bakre sannolikheten $P(e1e2|R)$ för varje diploid genotyp, och vinnaren kallas.

DRAGEN för IDPE Dx tillämpar automatisk filtrering. Se [VCF-filanteckningar i könscellsarbetsflödet på sidan 13](#) för ytterligare information.

DRAGEN Bestämning av somatisk variant

DRAGEN Somatic Small Variant Caller tar mappade och justerade DNA-avläsningar som indata och bestämmer SNV och indels genom lokal *de novo*-sammansättning av haplotyper i en aktiv region. För att aktivera DRAGEN Somatic Small Variant Caller, välj en somatisk variantapplikation.

Bestämning av somatisk variant används vanligtvis för tumörprover. Med detta arbetsflöde gör DRAGEN inga ploidinantaganden, vilket möjliggör detektering av lågfrekventa alleler. För loci med täckning upp till 100x i tumörprovet, har DRAGEN en detekteringströskel vid variantallelfrekvenser på 5 %. Gränsen skalas med ökande djup på en per-locus basis, och halveras varje gång täckningen fördubblas över 100x. Bestämbara referensregioner identifieras först med tillräcklig inriktningstäckning. Inom dessa referensregioner identifierar en skanning av de sorterade avläsningarna aktiva regioner, som är centrerade på pileup-kolumner med bevis på en variant i tumöravläsningarna. De aktiva regionerna är utfyllda med tillräckligt med sammanhang för att täcka betydande, icke-referensinnehåll i närheten. Om det finns bevis på indels får de aktiva regionerna ytterligare utfyllnad.

Justerade avläsningar klipps inom varje aktiv region och sätts samman till en De Bruijn-graf. Kanterna på de klippta avläsningarna viktas av observationsräkningar, med referenssekvensen som en ryggrad. Efter viss grafrensning och förenkling extraheras alla källa-till-sänkvägar som kandidathaplotyper. Varje haplotyp är Smith-Waterman anpassad till referensgenomet för att identifiera de varianter den representerar. För varje läs-haplotyp uppskattas sannolikheten $P(r|H)$ för att observera avläsningen med hjälp av en par-dold Markov-modell (HMM) under antagande att haplotypen är det sanna startprovet.

För att fastställa poäng för detektionsgräns för tumör (TLOD) skannar DRAGEN Somatic Small Variant Caller först enligt referensposition för varje somatisk kandidathändelse, samt referenshändelsen över den aktiva regionen. Den villkorade sannolikheten $P(r|e)$ för att observera varje överlappande avläsning uppskattas som det maximala $P(r|H)$ för haplotyper som stöder händelsen. Dessa kombineras till den villkorliga sannolikheten $P(r|E)$ för en händelsehypotes, E, som involverar en blandning av referens- och kandidatallelen över ett intervall av möjliga allelfrekvenser, och multipliceras för att ge den villkorade sannolikheten $P(R|E)$ att observera hela den lästa pileupen. Därifrån beräknas en TLOD-poäng som bevis på att en ALT-allel finns i tumörprovet på ett givet lokus.

DRAGEN for IDPE Dx tillämpar automatisk filtrering. Se [VCF-filanteckningar i det somatiska arbetsflödet på sidan 15](#) för ytterligare information.

Skapa en planerad körning

Använd följande steg för att ställa in en körning i Illumina Run Manager antingen på NextSeq 550Dx eller med en webbläsare på en nätverksansluten dator. Använd en webbläsare på en nätverksansluten dator om du vill importera exempeldata. Se Illumina Run Manager för NextSeq 550Dx programvaruguide (dokumentnr 200025239) för instruktioner om åtkomst av Illumina Run Manager från en nätverksansluten dator.

Det finns två olika sätt att skapa en ny, planerad körning:

- **Import Run** (Importera körning) – Använd ett provark från en befintlig körning som mall för en ny körning. Se Illumina Run Manager för NextSeq 550Dx programvaruguide (dokumentnr 200025239) för information om hur man importerar en körning.
- **Create Run** (Skapa körning) – Ange körningsparametrar manuellt. Följande instruktioner beskriver hur man skapar en körning.

OBS! De obligatoriska inmatningsfälten i användargränssnittet är markerade med en asterisk (*).

App

1. Välj **Create Run** (Skapa körning) på fliken Planned (Planerad) på skärmen Runs (Körningar).
2. Välj DRAGEN for Illumina DNA Prep with Enrichment Dx-appen och sedan **Next** (Nästa).

Körningsinställningar

1. På skärmen Run Settings (Körningsinställningar), ska du ange ett unikt körnamn. Körningsnamnet identifierar körningen från sekvensering till analys.
2. **[Valfritt]** Ange en körningsbeskrivning för att ytterligare identifiera körningen.
3. Välj indexadaptersats(er) som används under förberedelse av biblioteket.
4. Granska läslängden och modifiera vid behov. Read 1 (Avläsning 1) och Read 2 (Avläsning 2) har ett standardvärde på 151 cykler. Index 1 och Index 2 har ett fast värde på 10 cykler, och kan inte ändras.
5. **[Valfritt]** Ange ett biblioteksror-ID.
6. Välj **Next** (Nästa).

Provdata

Exempeldata inkluderar prov-ID, brunnposition (indexplattans brunnposition) och biblioteksnamn. Vid användning av index A&B inkluderar brunnposition även plattidentifiering.

Det finns två olika sätt att ange provdata:

- **Import Samples** (Importerera prover) – Använd en mallfil som kan laddas ned på skärmen Sample Data (Provdata).
- **Manually** (Manuellt) – Ange provdata direkt i tabellen på skärmen Sample Data (Provdata).

Importerera prover

Vid planering av en sekvenseringskörning med användning av en webbläsare på en dator i nätverket finns en mallfil (*.csv) tillgänglig för nedladdning på skärmen Sample Data (Provdata). Mallfilen är inte tillgänglig för nedladdning när du hämtar Illumina Run Manager via NextSeq 550Dx operativsystemets programvara. Gör så här för att mata in provdata med hjälp av funktionen Import Samples (Importerera prover).

OBS! Slutför stegen för Körningsinställningar innan du fortsätter.

1. Välj **Download Template** (Ladda ned mall) för att ladda ner en tom CSV-fil.
2. Från mallfilen ska du ange provdata och sedan spara filen. Biblioteksnamn är valfritt.

OBS! Vid användning av Index A&B, måste data för kolumn B inkludera både plattans och brunnens position (indexplattans brunnposition). Exempel: A-A01, A-A02, A-A03.

3. Välj **Import Samples** (Importerera prover) och bläddra till mallfilen som innehåller provdatainformation från föregående steg.
4. Välj **Open** (Öppna), **Proceed** (Fortsätt), och sedan **Next** (Nästa).

OBS! Ändring av Sample ID (Prov-ID) innan du väljer Next (Nästa) kan leda till ett fel. Undvik fel genom att slutföra inställningen av körningen innan du gör ändringar.

Ange prover manuellt

Använd tabellen på skärmen Sample Data (Provdata) för att ange provdata manuellt.

1. Ange ett unikt prov-ID i fältet Sample ID (Prov-ID).
2. Använd **Well Position** (Brunnsposition) (Index A eller index B) eller **Plate - Well Position** (Platta-brunnposition) (Index A&B) för att välja tillhörande index för proverna. Fälten i7 Index, Index 1, i5 Index och Index 2 fylls i automatiskt.
3. **[Valfritt]** Ange ett biblioteksnamn.
4. Lägg till rader och upprepa steg 1–3 efter behov tills alla prover har lagts till i tabellen. Du kan lägga till flera rader samtidigt genom att först ange antalet rader som ska läggas till och sedan välja ikonen +. Du kan också ta bort rader genom att markera rutan bredvid radnumret och sedan klicka på papperskorgikonen.
5. Välj **Next** (Nästa).

Analysinställningar

1. Välj önskat analysarbetsflöde:
 - FASTQ-generering
 - FASTQ- och VCF-generering för ett könscellsarbetsflöde (manifestfil krävs)
 - FASTQ- och VCF-generering för ett somatiskt arbetsflöde (manifestfil krävs)
2. **[Valfritt] Generera ORA-komprimerade FASTQ** är aktiverat som standard. FASTQ ORA-kompression komprimerar förlustfritt FASTQ-filer upp till 5x jämfört med fastq.gz. Avmarkera **Generate ORA compressed FASTQs** (Generera ORA-komprimerade FASTQ) om expanderade data (fastq.gz) föredras.
3. För könscells- och somatiska arbetsflöden krävs en manifestfil. Använd rullgardinsmenyn **Manifest File Selection** (Val av manifestfil) för att välja en manifestfil. Manifestet är en tabbavgränsad BED (*.bed) -fil, som anger namn och platser för riktade referensregioner. Se [Manifestfil på sidan 8](#) för ytterligare information.
4. **[Valfritt]**För somatiska arbetsflöden, använd rullgardinsmenyn **Noise File Selection** (Val av brusfil) för att välja en systematisk brusfil.
En BED-fil (*.bed.gz) med platsspecifik brusnivå kan specificeras för att filtrera bort systematiskt brus. Se [Brusfiltrering \(valfritt\) på sidan 9](#) för ytterligare information.
5. Välj **Next** (Nästa).

KÖR Granskning

1. På skärmen Review (Granskning) ska du granska informationen för Run Settings (Körinställningar), Sample Data (Provdatabas) och Analysis Settings (Analysinställningar).
2. Välj **Save** (Spara).
Körningen sparas på fliken Planned (Planerad) på skärmen Runs (Körningar).

Inställningar

Om du vill visa eller ändra inställningar för DRAGEN for IDPE Dx-appen, ska du först välja ikonen från huvudskärmen. Välj sedan appen du vill visa eller ändra. Ett administratörskonto krävs för att ändra inställningarna.

Konfiguration

Konfigurationsskärmen visar följande programinställningar:

- **Library Prep Kits** (Biblioteksförberedande sats) – Visar appens standardsats för biblioteksförberedelse. Denna inställning kan inte ändras.
- **Index Adapter Kits** (Indexadaptersatser) – Visar appens standardindexadaptersats. Denna inställning kan inte ändras.

- **Read lengths** (Läslängder) – Läslängder är inställda på 151 för appen som standard men kan ändras när körningen skapas.
- **Manifest and Noise Files** (Manifest- och brusfiler) – Ladda upp och ändra inställningar för manifest- och brusfiler.
 - Välj **Upload File** (Ladda upp fil) för att ladda upp filer för användning i analys.
 - Välj knappen **Default** (Standard) för att ställa in filen som standardmanifest eller brusfil som valts under körningen när applikationen väljs.
 - Markera **Enabled** (Aktiverad) för att ställa in filen så att den visas i rullgardinsmenyn under körningen.

Behörigheter

Använd kryssrutorna på skärmen Permissions (Behörigheter) för att hantera användaråtkomst för appen.

Manifestfil

Vid användning av DRAGEN for IDPE Dx krävs en manifestfil för följande arbetsflöden:

- FASTQ- och VCF-generering för ett könscellarbetsflöde
- FASTQ- och VCF-generering för ett somatiskt arbetsflöde

Manifestfilen är en tabbavgränsad textfil, som använder BED-formatet (*.bed), som anger namn och platser för riktade referensregioner. Huvuddelen av manifestfilen är regionavsnittet och bör innehålla följande datakolumner:

Kolumner	Beskrivning
Namn	Unikt, användarspecificerat namn för målet
Kromosom	Kromosomplats (t.ex. chr10, chr5, etc.)
Börja	1-baserat index för målets startposition
Stopp	1-baserat index för målets stopp-position
Uppströms sondlängd	Längden på uppströmssonden. För DRAGEN for IDPE Dx-appen ska detta ställas in på 0.
Nedströms sondlängd	Längden på nedströmssonden. För DRAGEN for IDPE Dx-appen ska detta ställas in på 0.

OBS! Ett giltigt manifestfilformat krävs för analys. DRAGEN stoppar analys om manifestfilen är ogiltig.

Brusfiltrering (valfritt)

Det systematiska brusfiltret är tillgängligt för bestämning av somatisk variant, och kan användas för att minska falskt positiva bestämningar genom att ta hänsyn till plats-specifikt brus. Den systematiska brusfilen genereras genom att först samla in ca 50 normala prover (helst specifika för panelen, biblioteksförberedelse och sekvenserare), och sedan delas summan av allelfrekvenser under 30 % på varje plats med tillräcklig täckning med det totala antalet prover (allelfrekvenser över 30 % förutsätts vara könscelesvarianter och inte brus). När brusvärdena väl har genererats, filtreras somatiska varianter som upptäcks på denna plats.

Filtret kan användas i tumör-normalt läge men är särskilt användbart för körningar med endast tumör där en matchad normal inte är tillgänglig. Den systematiska brusfilen måste använda en BED-fil, som har ett (* .bed.gz) tillägg och måste innehålla fyra kolumner: Kromosom, start, slut och plats-specifika brusnivåer för varje rad. Systematisk brusfiltrering är valfritt.

Analysutdata

Pågående körningar visas på fliken Active (Aktiv). Slutförda körningar visas på fliken Completed (Slutförd). DRAGEN for IDPE Dx skapar en unikt namngiven analysmapp för varje analys, som är separat från mappen som innehåller sekvenseringsdata. Analysmappen innehåller följande information:

- Använd manifestfil
- Programversion
- Prov-ID
- Totalt justerade läsningar
- Procent av justerade avläsningar per prov
- Antal bestämda SNV per prov
- Antal bestämda indels per prov
- Täckningsstatistik

Analysutdatafiler

Platsen för analysmappen anges med inställningen för External Storage for Analysis Results (Extern lagring för analysresultat). Se Illumina Run Manager för NextSeq 550Dx programvaruguide (dokumentnr 200025239) för ytterligare information om inställningen för External Storage for Analysis Results (Extern lagring för analysresultat).

På skärmen Run Details (Kördetaljer) visar fältet External Location (Extern plats) sökvägen för sekvenseringsdata. Det unika analysmappnamnet finns i fältet Analysis Output Folder (Analysutdatamapp) på skärmen Run Details (Kördetaljer). Exakt vilka filer som genereras beror på vilket analysarbetsflöde som används. Följande analysutdatafiler genereras av applikationen.

OBS! Om ett begränsningsfel för sökvägens maximala längd uppstår vid åtkomst av analysutdatafiler, kan du försöka flytta filen till en kortare sökvägsplats eller använda en annan metod för att öppna filen.

Utdatafil	Beskrivning
Variantsammanfattningsrapport (*.pdf)	Innehåller en sammanfattning av filinformationen, programversionerna, provinformationen, läsnivåstatistiken och SNV, insertioner, deletioner och täckningssammanfattningar. Endast könscells- och somatiska arbetsflöden producerar en variantrapport.
FASTQ (*.fastq.gz eller *.fastq.ora)	Intermediära filer som innehåller basbestämningar med kvalitetspoäng. FASTQ-filer är inpassningsstegets primära indata. När ORA-kompression väljs, används filtillägget *.fastq.ora.
Justering av BAM-filer (*.bam)	Innehåller justerade läsningar för ett givet prov.
Genome VCF-filer (*.gvcf.gz)	Innehåller genotypen för varje position, oavsett om den bestäms som en variant eller som en referens.
VCF-filer (*.vcf.gz)	Innehåller varianter bestämda vid varje position.
Statistikrapport för körning (*.csv)	Innehåller kvalitetsstatistik om körningen, inklusive icke-indexerad total avkastning och Q30-poäng.

FASTQ-filer

FASTQ (*.fastq.gz, *.fastq.ora) är ett textbaserat filformat som innehåller basbestämningar och kvalitetsvärden per läsning. Varje fil innehåller följande information:

- Providentifierare
- Sekvensen
- Phred-kvalitetspoäng i ett ASCII + 33-kodat format

Providentifieraren formateras enligt följande:

```
@Instrument:Körnings-ID:Flödescell-ID:Spår:Ruta:X:Y
Avläsningsnummer:Filterflagga:0:Provnummer
Exempel:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

BAM-filer

En BAM-fil (*.bam) är den komprimerade binära versionen av en SAM-fil (sequence alignment map) som används för att representera justerade sekvenser upp till 128 Mb. BAM-filer använder filnamnsformatet `SampleName_S#.bam`. # är provnumret som bestäms av den ordning som proverna listas för körningen. I multinodsläge är S# satt till S1, oavsett ordningen på samplet.

BAM-filer innehåller en rubriksektion och en inriktningssektion:

- **Header** (Rubrik) – Innehåller information om hela filen, som provnamn, provlängd och justeringsmetod. Justeringar i inriktningssektionen är associerade med specifik information i rubriksektionen.
- **Alignments** (Justeringar) – Innehåller läsnamn, lässekvens, läskvalitet, anpassningsinformation och anpassade taggar. Läsnamnet inkluderar kromosomen, startkoordinaten, justeringens kvalitet och matchningsbeskrivningssträngen.

Justeringssektionen innehåller följande information för varje läs- eller läspar:

- AS: Inriktningskvalitet med parad ände.
- RG: Läsgrupp, som anger antalet läsningar för ett specifikt prov.
- BC: Streckkodstagg, som indikerar det demultiplexade prov-ID som är kopplat till läsningen.
- SM: Ensidig inriktningskvalitet.
- XC: Matcha deskriptorsträng.
- XN: Tagg för amplikonnamn, som registrerar det amplikon-ID som är kopplat till läsningen.

BAM-indexfiler (*.bam.bai) innehåller en förteckning över motsvarande BAM-fil.

VCF-filer

Filer med variantbestämningsformat (*.vcf) innehåller information om varianter som finns på specifika positioner i ett referensgenom.

VCF-filhuvudet inkluderar VCF-filformatversionen och variantbestämningsversionen, och anger de anteckningar som används i resten av filen. VCF-huvudet inkluderar också referensgenomfilen och BAM-filen. Den sista raden i huvudet innehåller kolumnhuvudet för datalinjerna. Var och en av VCF-filens datarader innehåller information om en enda variant.

Tabell 1 VCF-filrubriker

Rubrik	Beskrivning
CHROM (Kromosom)	Referensgenomets kromosom. Kromosomerna visas i samma ordning som referens-FASTA-filen.

Rubrik	Beskrivning
POS	Variantens enkelbasposition i referenskromosomen. För enkelnukleotidvarianter (SNV) är denna position referensbasen med varianten. För indels är denna position referensbasen omedelbart före varianten.
ID	rs (referens SNP) numret för SNP erhållen från <code>dbSNP.txt</code> , om tillämpligt. Om flera rs-nummer finns på den här platsen avgränsas listan med semikolon. Om en dbSNP-post inte finns på denna position används en saknad-markör ('.').
REF	Referensgenotypen. Till exempel representeras en deletion av ett enda T som referens TT och alternativ T. En enkel nukleotidvariant A till T representeras som referens A och alternativ T.
ALT	Allelerna som skiljer sig från referensen läses. Till exempel representeras en insertion av ett enda T som referens A och alternativ AT. En enkelnukleotidvariant A till T representeras som referens A och alternativ T.
QUAL	En Phred-skalad kvalitetspoäng tilldelad av variantbestämaren. Högre poäng indikerar högre tilltro till varianten och lägre sannolikhet för fel. För kvalitetspoäng Q är den uppskattade sannolikheten för ett fel $10^{-(Q/10)}$. Till exempel har uppsättningen Q30-bestämningar en felfrekvens på 0,1 %. Många variantbestämare tilldelar kvalitetspoäng baserat på deras statistiska modeller, som är höga i förhållande till den observerade felfrekvensen.

Tabell 2 VCF-filanteckningar i könscellsarbetsflödet

Rubrik	Beskrivning
FILTER	<p>Om alla filter godkänns, anges PASS i filterkolumnen. Möjliga FILTER-poster inkluderar:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL – Tillämpas om SNP-variant QUAL-poäng inte når tröskeln • DRAGENIndelHardQUAL – Tillämpas om indel variant QUAL poäng inte når tröskeln • LowDepth – Platsen filtreras eftersom täckningsdjupet inte når tröskeln • LowGQ – Platsen filtreras eftersom genotypkvaliteten inte når tröskeln • PloidyConflict – Genotypbestämning från variantbestämmare överensstämmer inte med kromosomploidi • base_quality – Platsen filtreras eftersom medianbaskvaliteten för alt-läsningar på detta lokus inte når tröskeln • filtered_reads – Platsen filtreras eftersom en för stor del av läsningarna har filterats bort • fragment_length – Platsen filtreras eftersom den absoluta skillnaden mellan medianfragmentlängden för alt-läsningar och medianfragmentlängden för ref-läsningar på detta ställe överskrider tröskeln • low_depth – Platsen filtreras eftersom läsdjupet är för lågt • low_frac_info_reads – Platsen filtreras eftersom andelen informativa läsningar ligger under tröskeln • low_normal_depth – Platsen filtreras eftersom det normala provets läsdjup är för lågt • long_indel – Platsen filtreras eftersom indellängden är för lång • mapping_quality – Platsen filtreras eftersom medianmappningskvaliteten för alt-läsningar på detta lokus inte når tröskeln • multiallelic – Platsen filtreras eftersom mer än två alt-alleler passerar tumör-LOD • non_homref_normal – Platsen filtreras eftersom den normala provgenotypen inte är homozygot referens • no_reliable_supporting_read – Platsen filtreras eftersom det inte finns någon pålitlig stödjande somatisk läsning • panel_of_normals – Ses i minst ett prov i panelen av normaler vcf • read_position – Platsen filtreras eftersom medianen för avstånden mellan start/slut av läsning och detta lokus ligger under tröskeln • RMxNRepeatRegion – Platsen filtreras eftersom hela eller delar av variantallelen är en upprepning av referensen • strand_artifact Platsen filtreras på grund av allvarlig strängbias • str_contraction – Platsen filtrerad på grund av misstänkt PCR-fel där alt-allelen är en upprepning av referensen • too_few_supporting_reads – Platsen filtreras eftersom det finns för få stödjande avläsningar i tumörprovet • weak_evidence – Somatisk variantpoäng når inte tröskeln

Rubrik	Beskrivning
INFO	<p>Möjliga INFO-poster inkluderar:</p> <ul style="list-style-type: none"> • AC – Alleler i genotyper för varje ALT-allel, i samma ordning som anges. • AF – Allelfrekvens för varje ALT-allel, i samma ordning som anges. • AN – Det totala antalet alleler i kallade genotyper. • DB – dbSNP-medlemskap. • FS – Phred-skalat p-värde med Fishers exakta test för att upptäcka strängförspänning. • QD – Varianttillförlitlighet/-kvalitet efter djup. • R2_5P_bias – Poäng baserat på parningspartiskhet och avstånd från 5 primeslut. • SOR – Symmetrisk oddskvot på 2x2 beredskapstabell för att upptäcka strängförspänning. • DP – Ungefärligt läsdjup (informativt och icke-informativt); vissa läsningar kan ha filterats baserat på mapq etc. • END – Stopp position för intervallet. • FractionInformativeReads – Bråkdelen av informativa läsningar av det totala antalet läsningar. • MQ – RMS Kartläggningskvalitet. • MQRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs Ref läs mappningskvaliteter. • ReadPosRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs. Ref läs positionsbias. • SOMATIC – Minst en variant på denna position är somatisk.
FORMAT	<p>Formatkolumnen visar fält separerade med kolon. Till exempel GT:GQ. Tillgängliga fält inkluderar:</p> <ul style="list-style-type: none"> • AD – Alleldjup (räknas endast informativa läsningar av de totala läsningarna) för ref- och alt-alleler i den ordning som anges. • AF – Allelfraktioner för alt-alleler i den ordning som anges. • DP – Ungefärligt läsdjup (avläsningar med MQ=255 eller med dåliga kompisar filteras). • F1R2 – Antal avläsningar i F1R2-parorientering som stöder varje allel. • F2R1 – Antal avläsningar i F2R1-parorientering som stöder varje allel. • GT – Genotyp. 0 motsvarar referensbasen, 1 motsvarar den första posten i ALT-kolumnen och så vidare. Snedstreck (/) indikerar att ingen fasinformation är tillgänglig. • MB – Per-prov komponentstatistik för att upptäcka parningspartiskhet. • PS – Fysisk fas-ID-information, där varje unikt ID inom ett givet prov (men inte över prover) kopplar samman poster inom en fasningsgrupp. • SB – Per-prov komponentstatistik, som omfattar Fishers Exact Test för att upptäcka strängbias. • SQ – Somatisk kvalitet.

Rubrik	Beskrivning
PROV	Exempelkolumnen ger de värden som anges i FORMAT-kolumnen.

Tabell 3 VCF-filanteckningar i det somatiska arbetsflödet

Rubrik	Beskrivning
FILTER	<p>Om alla filter godkänns, anges PASS i filterkolumnen. Möjliga FILTER-poster inkluderar:</p> <ul style="list-style-type: none"> • base_quality – Platsen filtreras eftersom medianbaskvaliteten för alt-läsningar på detta lokus inte når tröskeln • filtered_reads – Platsen filtreras eftersom en för stor del av läsningarna har filtrerats bort • fragment_length – Platsen filtreras eftersom den absoluta skillnaden mellan medianfragmentlängden för alt-läsningar och medianfragmentlängden för ref-läsningar på detta ställe överskrider tröskeln • low_depth – Platsen filtreras eftersom läsdjupet är för lågt • low_frac_info_reads – Platsen filtreras eftersom andelen informativa läsningar ligger under tröskeln • low_normal_depth – Platsen filtreras eftersom det normala provets läsdjup är för lågt • long_indel – Platsen filtreras eftersom indellängden är för lång • mapping_quality – Platsen filtreras eftersom medianmappningskvaliteten för alt-läsningar på detta lokus inte når tröskeln • multiallelic – Platsen filtreras eftersom mer än två alt-alleler passerar tumör-LOD • non_homref_normal – Platsen filtreras eftersom den normala provgenotypen inte är homozygot referens • no_reliable_supporting_read – Platsen filtreras eftersom det inte finns någon pålitlig stödjande somatisk läsning • panel_of_normals – Ses i minst ett prov i panelen av normaler vcf • read_position – Platsen filtreras eftersom medianen för avstånden mellan start/slut av läsning och detta lokus ligger under tröskeln • RMxNRepeatRegion – Platsen filtreras eftersom hela eller delar av variantallelen är en upprepning av referensen • strand_artifact Platsen filtreras på grund av allvarlig strängbias • str_contraction – Platsen filtrerad på grund av misstänkt PCR-fel där alt-allelen är en upprepning mindre än referensen • too_few_supporting_reads – Platsen filtreras eftersom det finns för få stödjande avläsningar i tumörprovet • weak_evidence – Somatisk variantpoäng når inte tröskeln • systematic_noise – Platsen filtreras baserat på tecken på systematiskt brus i normala

Rubrik	Beskrivning
INFO	<p>Möjliga INFO-poster inkluderar:</p> <ul style="list-style-type: none"> • DP – Ungefärligt läsdjup (informativt och icke-informativt); vissa läsningar kan ha filterats baserat på mapq etc. • END – Stopp position för intervallet. • FractionInformativeReads – Bråkdelen av informativa läsningar av det totala antalet läsningar. • MQ – RMS Kartläggningskvalitet. • MQRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs Ref läs mappningskvaliteter. • ReadPosRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs. Ref läs positionsbias. • AQ – Systematisk bruspoäng. • hotspot – Känd somatisk sida, används för att öka bestämmningens tillförlitlighet. • SOMATIC – Minst en variant på denna position är somatisk.
FORMAT	<p>Formatkolumnen visar fält separerade med kolon. Till exempel GT:GQ. Tillgängliga fält inkluderar:</p> <ul style="list-style-type: none"> • AD – Alleldjup (räknas endast informativa läsningar av de totala läsningarna) för ref- och alt-allelerna i den ordning som anges. • AF – Allelfraktioner för alt-alleler i den ordning som anges. • DP – Ungefärligt läsdjup (avläsningar med MQ=255 eller med dåliga kompisar filteras). • F1R2 – Antal avläsningar i F1R2-parorientering som stöder varje allel. • F2R1 – Antal avläsningar i F2R1-parorientering som stöder varje allel. • GP – Phred-skalade posteriora sannolikheter för genotyper enligt definitionen i VCF-specifikationen. • GQ – Genotypkvalitet. • GT – Genotyp. 0 motsvarar referensbasen, 1 motsvarar den första posten i ALT-kolumnen och så vidare. Snedstreck (/) indikerar att ingen fasinformation är tillgänglig. • MB – Per-prov komponentstatistik för att upptäcka parningspartiskhet. • PL – Normaliserade, Phred-skalade sannolikheter för genotyper enligt definitionen i VCF-specifikationen. • PRI – Phred-skalade tidigare sannolikheter för genotyper. • PS – Fysisk fas-ID-information, där varje unikt ID inom ett givet prov (men inte över prover) kopplar samman poster inom en fasningsgrupp. • SB – Per-prov komponentstatistik, som omfattar Fishers Exact Test för att upptäcka strängbias. • SQ – Somatisk kvalitet.
PROV	Exempelkolumnen ger de värden som anges i FORMAT-kolumnen.

Genome VCF-filer

Genome VCF-filer (*.gvcf.gz) följer en uppsättning konventioner för att representera alla platser inom genomet i ett relativt kompakt format. gVCF-filerna inkluderar alla platser inom området av intresse i en enda fil för varje prov. gVCF-filen visar no-calls på positioner som inte klarar alla filter. En genotyp-tag (GT) av ./ indikerar ett no-call.

Repetera analys

Du kan repetera en analys om analysen stoppades, misslyckades eller om du vill analysera om en körning med andra inställningar. Gör så här för att repetera analys:

1. På skärmen Run (Kör) ska du välja fliken Completed (Slutförd) och sedan det körningsnamn som ska oanalyseras.
Om Repetera analys utfördes tidigare, ska du välja körningsnamnet för den Överordnade körningen.
2. Från skärmen Run Details (Kördetaljer), efter Sequencing Information (Sekvenseringsinformation), ska du välja **Requeue Analysis** (Repetera analys).
3. Välj ett alternativ:
 - Requeue analysis with no changes (Repetera analys utan några ändringar)
 - Edit run settings and requeue analysis (Redigera körningsinställningar och repetera analys)
 - Requeue analysis with a different application (Repetera analys med en annan app)
4. Bekräfta att platsen där sekvenseringsdata för närvarande finns, finns i fältet **Sequencing data file path** (Sökväg för sekvenseringsdata).

OBS! Sökvägen till sekvenseringsdata bör matcha sökvägen i inställningen för External Storage for Analysis Results (Extern lagring för analysresultat). Se Illumina Run Manager för NextSeq 550Dx programvaruguide (dokumentnr 200025239) för information om hur man ändrar den externa lagringsvägen.

5. Ange en orsak till oanalys.
6. Välj **Requeue Analysis** (Requeue Analysis).
7. Redigera de önskade ändringarna av Run Settings (Körinställningar), Sample Data (Provdata) och Analysis Settings (Analysinställningar).
8. Välj **Save** (Spara). Analysen påbörjas med de aktuella analysparametrarna.

Teknisk hjälp

För teknisk hjälp, kontakta Illumina Teknisk Support.

Webbplats: www.illumina.com

e-post: techsupport@illumina.com

Säkerhetsdatablad (SDS) – Finns på Illuminas webbplats på support.illumina.com/sds.html.

Produktdokumentation – Kan hämtas på support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 USA
+1.800.809.ILMN (4566)
+1 858 202 4566 (utanför Nordamerika)
techsupport@illumina.com
www.illumina.com



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

FÖR IN VITRO-DIAGNOSTISKT BRUK.

© 2023 Illumina, Inc. Med ensamrätt.

illumina[®]