

NGS – Teknisk vägledning

Användningen av storskalig sekvensering med andra och tredje generationens instrument (s.k. NGS, next-generation sequencing alternativt MPS, massivt parallell sekvensering) i klinisk drift skiljer sig inte principiellt från andra sekvenseringstekniker, såsom Sangersekvensering. Vid användandet av NGS/MPS finns det dock fler tekniska aspekter som ska beaktas. Nya metoder och uppdateringar lanseras ofta. Validering och verifiering av metoder som används inom det svenska ackrediteringsförfarandet ger ett fullgott stöd för att säkerställa god klinisk kvalitet. Det är viktigt att beakta att hela kedjan, från DNA-preparation till bioinformatisk behandling påverkar resultatet - det beslutsunderlag som ges för variantbedömning. En validering tar hänsyn till alla steg, inklusive exempelvis instrument, biblioteksberedning, anrikningsreagens, sekvenseringsreagens, programvara för bas- och variantbestämning och val av referensgenom för sekvensinpassning.

ORDLISTA

basbestämning, *eng. base call*, att bioinformatiskt avgöra vilken bassekvens som genererar rådata. Även "basavrop".

bibliotek, *eng. library*, – flera olika molekyler prov-DNA sönderdelat i lagom fragment, försedda med för sekvensering nödvändiga artificiella oligonukleotider.

biblioteksberedning, *eng. library preparation*. Att färdigställa ett bibliotek från prov-DNA. Även "bibliotekspreparation".

demultiplexning, *eng. demultiplexing*, bioinformatiskt avkodande av molekyllär kod som biblioteket försetts med och utskiljande av vilka läsningar i en körning som hör till vilket bibliotek.

infångning, *eng. capture*, biokemisk/fysisk selektion av delar av bibliotek

inmärkning, *eng. label*, artificiell spårbar förändring av biomolekyl som t ex kan läsas med instrument.

läsning, *eng. read*, sekvens från ett fragment ur ett bibliotek.

obestämda läsningar, *eng. undetermined reads*, läsningar med otydlig molekyllär kod som inte med säkerhet kan bestämmas till ett givet i körningen laddat bibliotek.

sekvensinpassning, *eng. alignment*, att bioinformatiskt bestämma troligaste genomiskt ursprung till en läsning. Även "inpassning", "mappning".

splitsning, *eng. splicing*, processen där introner avlägsnas ur pre-RNA och exoner fogas samman till en kortare mogen mRNA-molekyl.

variantbedömning, *eng. variant interpretation*, manuell sammanfattande tolkning av en variant i relation till patientens symtom (även "tolkning" eller *eng. "variant triage"*).

variantbestämning, *eng. variant call*, att bioinformatiskt avgöra vilken genomisk variant på en specifik position som genererade de läsningar och den inpassning som ses. Även "variantavrop".

Sekvenseringsteknologi

De vanligaste storskaliga instrumenten som saluförs av Illumina bygger precis som vid Sangersekvensering på elongering med inmärkte nukleotider för optisk avläsning. Dessa är dock reversibelt terminerade. Andra generationens instrument från Thermo-Fisher använder istället oterminerad elongering med efterföljande pH-mätning. Tredje generationens instrument från Pacific-Biosciences och Oxford Nanopore använder sig av optisk respektive elektronisk mätning/analys.

Sekvenseringens syfte

Syftet med DNA-sekvensbestämning kan antingen vara:

- Att bekräfta eller utesluta närvaro av en specifik sekvensvariant (genotypning/anlagsbärartestning) eller
- Att förutsättningslöst söka efter avvikelser från en referenssekvens (screening).

Dessa båda tillämpningar ställer olika kvalitetskrav på utförandet av analysen.

Vilket sekvenseringsdjup som bör användas beror på typ av variation som finns och förväntas i sekvensen samt dess kopietal. Denna parameter och trösklar för alleliska procentsatser bör därför bestämmas empiriskt och valideras.

Metodsteg sekvensering & sekvensanalys

Patientmaterial

Utgångsmaterial för DNA-sekvensbestämning kan vara patientprov som hanteras i enlighet med för laboratoriet fastställda kvalitetskrav. Olika vävnader kan ge svar på olika frågor, och kräver olika behandling. Generellt gäller att vävnadsval är en del av hela processen som behöver valideras. Blodprov är väl beprövat för att söka konstitutionella förändringar. Ett salivprov kan innehålla större mängder DNA från andra organismer, vilket påverkar resultatet av helgenomsekvensering. Detta gäller i mindre mån kindsrap. Hudbiopsier fungerar normalt väl. Kvaliteten på DNA från FFPE material varierar med metod. Rent generellt ses kortare fragment, och särskilda extraktions- och infångningsmetoder kan behövas.

DNA-preparation

Kvalitet på DNA extraktion kan ha en stor inverkan på sekvenseringsresultatet. Det rekommenderas att laboratoriet sätter upp krav för godkänd DNA kvalitet och kvantitet, t ex godkänd kvot för A260/A280. Koncentrationsbestämning bör i den mån möjligt genomföras med en fluorescensbaserad metod. Fler steg i den experimentella processen nedströms medför ökad känslighet för olika former av kontaminationer. Olika biblioteksberednings- och sekvenseringstekniker kan vara känsliga för olika problem med DNA-kvalitet. Man bör även beakta elueringsbuffertens sammansättning, då t ex vissa metoder för biblioteksberedning är känsliga för EDTA. Särskilt för tekniker som arbetar med långa fragment kan försiktig hantering behövas.

Biblioteksberedning

Kvaliteten på beredda bibliotek inverkar på både täckning och sekvenseringskvalitet och kontrolleras lämpligen. Vid kontroll av kvaliteten är det önskvärt att utesluta artefakter från t ex adaptor-dimerer. Dessutom kontrolleras att biblioteket innehåller fragment av korrekt storlek.

Infångning / Urval

För att fånga upp den önskade regionen finns många olika tekniker man kan använda sig av. Dessa tekniker skiljer sig i sättet den genomiska regionen fångas in/amplifieras och detta har konsekvenser för hur djupt man bör sekvensera, vilken täckning man kan kräva på den enskilda positionen och vilka sekvenserings- eller amplifieringsartefakter/problem som kan uppkomma. Ett exempel på detta är tekniker som bygger på amplifiering med multiplex PCR, som kan ge upphov till allelförlust (allelic dropout) av amplicon, och sedan ger sekvensering av amplicon som till viss del kan komma från

samma DNA-molekyl. Detta måste man ta hänsyn till när man bestämmer vilket täckningsdjup man bör ha, samt om överlapp av primers behövs. Andra tekniker bygger på infångning med prober, med påföljande PCR amplifiering. Även detta ger upphov till PCR-duplikat, dock kan dessa lättare identifieras och avlägsnas i den bioinformatiska analysen.

Vissa varianter kan undgå detektion om sekvensbestämningen utförs endast på en DNA-sträng. Vid amplifieringsbaserad sekvensbestämning av prover bör sekvensering utföras på båda DNA-strängarna, d.v.s. med sekvensering i båda riktningarna. Sekvensspecifika artefakter förekommer. Det rekommenderas att söka överlappande sekvenser med olika startställen. För metoder med hög felfrekvens per bas är ett flertal sådana att föredra.

Vid alla typer av infångning bör man beakta att vissa sekvenser sannolikt är dåligt representerade i det slutgiltiga biblioteket. Detta gäller t ex regioner med hög andel GC.

Klustring [Illumina-specifikt]

För att sekvenseringen ska bli så korrekt som möjligt är det viktigt att inmängden optimeras så att klustertätheten hamnar inom instrumentets specifikation. Både vid för låg och för hög klustertäthet kommer antalet användbara läsningar bli lågt. För vissa Illumina-instrument (HiSeq 3000, HiSeq 4000, HiSeq X) kommer för låg klustertäthet dessutom ge upphov till duplikat på grund av 'pad hopping'.

Sekvensbestämning

Efter sekvensering bör man kontrollera att körningen motsvarar de krav man ställt upp för att sekvenseringens kvalitet. Några viktiga parametrar är andelen baser med en viss kvalitet (t.ex. % >Q30), färgbalans, andelen läsningar som inte säkert tillhör de körda proven (andelen obestämda läsningar) efter demultiplexning och att den effektiva högkvalitativa läslängden är så lång som önskat.

Basbestämning

Att tolka rådata till sekvens kräver program med stor kännedom om den specifika sekvenseringsmetoden. Ofta tillhandahåller instrumentleverantören acceptabel programvara. Dessa genererar i regel felsannolikhetsuppskattningar per bas som sedan används av programvara nedströms. Viss försiktighet bör iaktas med att tolka dessa uppskattningar som absoluta storheter, men relativt andra baser i samma eller andra experiment fungerar de i regel väl.

Sekvensinpassning

Val av referensgenom påverkar sekvensinpassning och efterföljande variantbestämning. S.k. decoy-sekvenser, såsom vanliga virus, och vanlig humansekvens som inte ingår i referensgenomet (t ex från JC Venter eller HuRef) ger färre inpassningsfel och därmed färre variantbestämningsfel [Li et al]. Teoretiskt är det fördelaktigt att vid sekvensinpassning använda en etnisk referens, d.v.s. ett referensgenom som väl återspeglar individens genetiska bakgrund. I klinisk praxis används detta sällan. Patientens genetiska bakgrund är troligen inte helt känd - eller tillgänglig som referensgenom. Det kan vara opraktiskt vid variantannotering då bra annoteringsdatabaser saknas. Val av inpassningsalgoritm och parametrar för denna har stor påverkan på slutresultatet. Kännedom om felprofiler i läsningar och kvalitetsvärden, samt tekniska vägval gällande t ex läslängd, gör dessa plattformsspecifika. bwa, Mosaik, NovoAlign och CLC är vanliga. De flesta program ger en

sannolikhetsbedömning för att en läsning placerats korrekt, t ex i form av ett värde på inpassningskvalitet.

Variantbestämning

Med variantbestämning menas processen då avvikelser från referensgenomet bestäms. Detta sker normalt genom att baser i olika läsningar som inpassats mot samma position i referensgenomet kontrolleras och sannolikheten för en alternativ allel på positionen bedöms, t ex genom en majoritetsomröstning. Sannolikheten att varianten är korrekt bestämd skattas också, som en funktion av baskvalitet och inpassningskvalitet hos de olika läsningarna av positionen. Inpassningen görs normalt parvis med en läsning, eller ett läspar, mot referensgenomet. Detta och en god skattning av variantbestämningskvaliteten kan vara tillräcklig. En fullständig upplösning av den multipla inpassningen av alla läsningar mot alla är resurskrävande, men görs i en del pipelines delvis genom s.k. realignment (återinpassning) lokalt för positioner där insertioner eller deletioner ropats av. Manuell bedömning av områden där mjukvara har svårigheter med tolkning skall utföras med största försiktighet. Tolkning av sekvensdata bör utföras av mer än en behörig person. Avvikelser från referenssekvens bedömda som sjukdomsassocierade som påvisas vid screening bör verifieras. I det fall avvikelserna har detekterats med metodik som endast validerats för screening bör den verifieras med en alternativ metod, gärna en som validerats för anlagstest. Fynd med en sådan metod verifieras genom test med nytt templat. Vanliga strategier för verifikation innefattar anlagstest i familjen, omtest med nytt patientprov, oanvänd alikvot, ny amplifiering - med samma eller hellre andra primers.

Variantannotering

Processen där bestämda varianter kopplas med för varianten tillgänglig information. T ex bör slagning i större frekvensdatabaser, såsom ExAC, 1000 genomes, 6500 ESP och nationella frekvensdatabaser göras för att få en god uppfattning om variantens vanlighet. Det rekommenderas även att lokala databaser över varianter tillfrågas, för att möjliggöra filtrering av en stor mängd lokalt förekommande tekniska artefakter. Detta är av stor vikt för att ge tillräcklig specificitet vid stora paneler i avsaknad av normalprov. Mjukvara för prediktion av funktionsförändring hos protein, såsom till exempel SIFT, Polyphen, CADD bör rådfrågas, liksom dito för splitsningsställen, såsom SPIDEX. Giltiga arvmönster noteras i relation till släktanamnes, och jämförs om möjligt med kända arvmönster för sjukdomsassocierade varianter i aktuell gen. Databaser med bedömda varianter i relation till fenotyp konsulteras: allmänna såsom ClinVar och HGMD samt sjukdomsspecifika såsom ENIGMA och InSiGHT.

Variantbedömning

Vid bedömning av om en sekvensvariant förklarar patientens symtom bör man följa uppsatta kriterier för att få en objektiv bedömning. Nuvarande rekommendation är att följa ACMGs riktlinjer (se SFMGs rekommendationer för Provrapportens utformning).

Datahantering

Provet skall vara spårbart under hela analysprocessen. Resultat och rapporter skall vara spårbara. För MPS skall identifierade genomiska varianter sparas, med dokumentation av mjukvara samt parametrar och kvalitetskontrolldata. Rådata och/eller läsningar bör sparas enligt laboratoriets aktuella rutiner, men åtminstone till dess slutsvar är distribuerat. Resultat skall jämföras med

dokumenterad referenssekvens och resultatet (d.v.s. identifierad variant eller konfirmerad normal sekvens) skall dokumenteras tillsammans med signatur och datum av behörig person som utför analysen.

Metodvalidering

Samtliga metoder för DNA-sekvensering som används inom klinisk diagnostik skall vara validerade. Syftet med valideringen är att fastställa tillförlitligheten för ett test, d v s att metoden mäter det man har för avsikt att mäta. Därför måste valideringen omfatta hela testprocessen "från ax till limpa" och inkludera alla steg i analysmetoden; från DNA extraktion, framställande av templat för sekvensering, sekvensering, bioinformatik/programvara som används för analys av resultat, eventuell konfirmering av sekvensvarianter samt rapportering av resultat. Många laboratorier omfattar inte alla steg lokalt. Senare steg kan valideras med väl definierade krav på tidigare steg, t ex provkrav eller certifieringskrav.

Varje sekvenseringsteknologi har styrkor och svagheter och de olika bioinformatiska verktygen måste beakta dessa egenskaper. Till exempel, varianter inom homopolymera regioner bör undersökas extra noggrant vid pyrosekvensering och halvledarsekvensering, medan tvåfärgssekvensering genom hybridiseringstekniker kräver specifika förfaranden för att separera färgerna. Laboratoriet bör undersöka sensitivitet och specificitet och andra kvalitetsparametrar för olika typer av sekvensvarianter. En intern databas som innehåller alla relevanta varianter utgör ett viktigt redskap för att identifiera plattformsspecifika artefakter.

Prestandan av ett diagnostiskt test baserat på NGS måste utvärderas med avseende på precision, analytisk sensitivitet och specificitet. Detta får dock precis som för andra tekniker göras under beaktande av metodens begränsningar avseende typen av varianter som kan hittas. Noggrannheten korrelerar med många kvalitetsmått, vilket kan utnyttjas vid kvalitetskontroll.

Precision hänvisar till överensstämmelse mellan upprepade mätningar av samma material. Ett tillräckligt antal prover (rekommenderat är minst 3) bör analyseras för att fastställa precisionen genom att bedöma reproducerbarhet (precision mellan körningar) och repeterbarhet (precision inom en och samma körning) under valideringstestet. Med reproducerbarhet menas att resultaten blir konsekventa från samma prov under olika villkor, såsom mellan olika körningar, olika provpreparat, med olika tekniker och med hjälp av olika instrument. En samstämmighet mellan 95-98% kan vara en bra riktlinje (Rehm et al. 2013).

Plattformvalidering fastställer att sekvenseringstekniken korrekt kan läsa en DNA-sekvens. Valideringen bör också undersöka hur väl olika typer av sekvensvarianter kan detekteras. För NGS-baserade diagnostiska analyser bör noggrannhet (överensstämmelse med förväntade resultat), analytisk precision (resultatets upprepbarhet), analytisk sensitivitet, specificitet och rapporterbart testområde bestämmas empiriskt under valideringen (ref Gargis et al. 2012).

För testets sensitivitet spelar det ingen roll om en variant missas på grund av sekvenseringstekniken eller den bioinformatiska analysen. Det spelar inte heller någon roll om det är analysen som sorterar bort falskt positiva resultat eller om de tas bort vid konfirmering. Analytisk specificitet och sensitivitet bör bestämmas under utvärderingen av den bioinformatiska pipelinen. Ur praktisk

synvinkel bör valideringen fokusera på genomiska regioner under utredning (regions of interest, ROI). Valideringstestet skall visa förmågan att upptäcka varianter i ROI som definierats under utvecklingen av analysen.

Kvalitetskontroll

Syftet med kvalitetskontroll är att visa att tillförlitligheten under drift inte avviker från den som fastställts under valideringen. Varje del i testet bör ha definierade kvalitetsmått och deras normala "range" (värdemängd) ska uppskattas vid valideringen. Test med kvalitetsmått som faller utanför den validerade värdemängden skall undersökas som indikatorer på problem med testmetoden. Laboratorierna uppmuntras att delta i kvalitetssäkringsprogram när testet har validerats och medverka i extern kvalitetsbedömning. I detta sammanhang uppmanas även laboratorier att dela väl karakteriserade prover och datafiler med varandra för att förbättra och standardisera praxis för diagnostik.

Det diagnostiska laboratoriet har att följa upp relevanta kvalitetsmått för (i) plattformen, (ii) alla analyser, (iii) alla prover. Alla kvalitetsmått som används i diagnostik bör noggrant beskrivas. Kvaliteten på en provanalys beror på många parametrar: t ex mängden data som producerats, andelen kluster för varje prov (vid multiplexning), andelen PCR-duplikat och täckningsdjup måste beaktas.

Plattformsspecifika kvalitetsparametrar noteras vid validering för att senare användas som kriterier för godkänt test.

Uppgifter om beräkning av kvalitetsmått skall vara väl dokumenterade för att göra tolkningen lätt. För att underlätta automatiserad hantering av kvalitetskontrollvärden (QC-värden) ska kvalitetsmått definieras och dokumenteras i en enhetlig terminologi och standardiserade filformat.

Alla sekvenseringskörningar måste övervakas för att säkerställa att man ligger inom instrumentets specifikation. Utöver detta bör körningsspecifika egenskaper övervakas kontinuerligt. Minimikrav bör etableras för viktiga kvalitetsmått, och viktiga analys- och provspecifika egenskaper sparas. Övervakningsdata ska inte redovisas utan användas som en kontinuerlig kvalitetsgranskning. Det är viktigt att hålla reda på undantag, såsom antalet gånger som ett prov har sekvenserats om för att nå fastställda kvalitetskriterier och korrigering av eventuella provförväxlingar.

Särskilt vid större provflöden bör en provspårningsmetod användas, eftersom NGS-arbetsflöden är komplexa och innefattar flera processteg både i labbet och i den bioinformatiska analysen. Att tidigt i kedjan avskilja en mängd prov som kan användas för kontroll mot provförväxling, eller för framställande av templat för en komplementär teknik, kan vara till hjälp.

Andelen icke-inpassade sekvensläsningar och sekvensläsningar utan index bör också följas eftersom det kan bidra till att identifiera grovt avvikande prover/analyser (på grund av kontaminering under arbetsflödet). Detsamma gäller kraftigt avvikande antal homozygota och heterozygota varianter för ett prov.

Slutligen bör jämförelser och uppföljning mellan olika analyser göras genom att inkludera generiska kvalitetsregioner. Beräkning av antalet avvikande bas-läsningar (icke-vildtypsvarianter), ogiltiga baser (betecknad som bas 'N') och sporadiska InDels i dessa regioner kan bidra till att identifiera avvikande prover. Uppföljning av dessa parametrar möjliggör dessutom jämförelse av olika versioner

av ett diagnostiskt test samt för jämförelser mellan prov. Olika sekvensplattformer, anrikningsmetoder, etc. kan också jämföras.

Kvalitetsmått

Teknikerna inom storskalig sekvensering förändras snabbt. Här belyser vi kvalitetsmått som idag anses gångbara som kvalitetskorrelat. Dessa etableras vid validering och verifiering, och följs sedan vid enskilda analyser. Flera av dem är användbara både för att först övergripande bedöma om ett prov har uppnått tillräcklig kvalitet för att tolkning ska vara meningsfullt, och som ett riktmärke för om en enskild variant sannolikt utgör en artefakt eller har tillräcklig teknisk kvalitet för att vara intressant för vidare bedömning.

- Klustringsdensitet [Klustring]

Illuminas teknik använder brygg-PCR på ett fast stöd för att generera tillräckligt mycket templat i lokalt unika fläckar för tillförlitlig optisk sekvensbestämning. Här är särskilt den s.k. klusterdensiteten av vikt. Denna ger en indikation om hur väl en sekvensering skulle falla ut redan på klustringsstadiet, och ett korrelat till den övergripande sekvenseringskvaliteten väl värt att optimera. Måttet har delvis sin motsvarighet i mängden laddat prov på ett Torrent-instrument, eller mängden bibliotek vid ett Nanopore-experiment.

- Fördelning av kluster per prov [biblioteksberedning, klustring, demultiplexning]

Beredda bibliotek bör kontrolleras avseende kvalitet. Noggrann kvantifiering samt bestämning av ingående fragmentlängder krävs för optimal pooling/multiplexning samt för att uppnå en klusterdensitet inom specificerat intervall. Förutsatt jämn pooling/multiplexning av upparbetade bibliotek bör antalet sekvensläsningar för respektive prov efter demultiplexning vara inom samma storleksordning. Kontroll av fördelningen av sekvensläsningar ger möjlighet att upptäcka problem relaterade till indexering (t ex felaktigt val av index vid biblioteksberedning eller felaktigt angiven indexering vid sekvensering).

- Baskvalitet [Sekvensering]

Kvaliteten över en körning kan påverkas av ett stort antal faktorer. Den reversibla terminatorns kvalitet är av stor vikt för att undvika fasfel i sekvenseringsfasen. Fasfel leder till sekvenser av låg tillförlitlighet, och i praktiken till kortare användbar läslängd än avsett. Sekvenser med låg kvalitet mot änden kan ibland bioinformatiskt trimmas ner till en längd med användbar kvalitet. Om inpassningsprogramvaran (t ex bwa eller Mosaik) är medveten om kvalitetsvärden är detta inte nödvändigt eller alltid ens önskvärt. Instrumenttillverkarna och fristående mjukvarutillverkare tillhandahåller program som genererar kvalitetsvärden (Q), typiskt på samma skala som Sanger-sekvenseringens Phred, med felsannolikheten skattad som $10^{-(Q/10)}$. Andel baser över den relativt stränga Q30 (1 fel på 1000 basbestämningar) används ofta som mått för en hel körning, och kan ge tydlig indikation på om det är värt att ta data vidare till nästa analyssteg. Baskvalitet sett per position i läsningen kan ge en indikation om hela körningen och vara användbart vid felsökning.

- Vertikal täckning [Hela processen]

Genomsnittlig täckningsgrad (i mål), andel baser (av mål) täckta till tröskelvärde – såsom 20x,

andel transkript med baser som är helt fullständiga – d.v.s. helt saknar baser med täckning under ett visst tröskelvärde – såsom 10x. Ibland används termen informativ täckning för vertikal eller horisontell täckning, då duplicerade läsningar och läsningar eller enskilda baser under något visst tröskelvärde i kvalitet utelämnats. Vid bedömning av täckningsdjup bör kvalitetskriterier baseras på informativ täckning i stället för rå täckning (Weiss, Van der et al. Zwaag 2013). Till exempel så kan gener med pseudogener eller repetitiva element visa hög täckning men ha låg informativ täckning om läsningar med dålig inpassningskvalitet tas bort.

- Horisontell täckning, Fullständighet, Läsningar i målet, Likformighet [Hela processen]

Ibland önskar man inte sekvensera hela genom, utan bara ett specifikt målområde. Andelen läsningar inom mål blir då, givet en fix sekvensmängd, en för kvalitet viktig parameter (“% on target”). Likformighet mellan regioner är på samma sätt viktigt, särskilt vid infångning eller amplifieringsbaserade metoder. Horisontell täckning, t ex uttryckt som andelen av baser i målen som uppnått en given kvalitet är ett viktigt mått för den kliniska prestandan av ett test, tillsammans med fullständighet; andelen måltranskript som uppnått 100 % horisontell täckning.

- Dupliceringsgrad [Biblioteksberedning, Klustring, Sekvensering]

Vid amplifiering, som t ex ofta behövs vid infångning i panel eller exomsekvensering, kan PCR-duplikat uppstå. Dessa är problematiska på flera sätt: dels leder de till en ojämn representation av täckning. Täckningen på respektive allel kan också bli ojämn. Detta kan leda till att en faktisk variant missas genom exklusion i kvalitetsfilter som kräver en jämn läsfördelning på båda alleler, och i båda riktningarna. Om fel i sekvensen introduceras tidigt och sedan amplifieras kan de statistiska modellerna som används vid variantbestämning bli lurade att en falsk variant finns.

Liknande fenomen kan uppstå optiskt vid avläsning, s.k. optiska duplikat, och även vid klustring på Illumina-plattformen - som är en form av bryggamplifiering. Duplikat kan i viss mån filtreras bort informatiskt, men hög dupliceringsgrad tyder på ett problem i processen.

- Andelen inpassade sekvensläsningar [DNA-preparation, Bibliotekspreparation, Sekvensering]

En lägre andel inpassade läsningar än normalt kan t ex indikera kontamination av provet, fel i demultiplexning eller problem vid sekvenseringen. Vid analys av salivprov med helgenomsekvensering är det normalt att 5-20% av läsningarna inte inpassats mot humana referensgenomet.

- GC-innehåll [Hela processen]

Det är i PCR-baserade metoder svårundvikligt, men inte önskvärt att sekvens med olika bassammansättning representeras olika väl. Mått på skevheten i GC-innehåll kan med fördel följas.

- Målets tillgänglighet [Bibliotekspreparation, Sekvensering, Analys]

Amplifieringstekniker och korta läsningar är begränsade i sin möjlighet att identifiera sekvens som inte är unik till läslängd och under rådande felfrekvens. Detta innebär praktiska problem för gener i genfamiljer och med snarlika pseudogener eller lågkomplexa delar, samt för gener som av andra anledningar har icke-unik sekvens. Index som sammanfattar möjligheten att nå enskilda baser med en given sekvenseringsteknologi finns ofta att tillgå, och kan etableras teoretiskt eller empiriskt för ett givet målområde. Ofta tillåter teknologier med längre läslängd, parade läsningar eller läsningar

knutna med en unik molekylär kod korrekt avläsning i en större del av arvsmassan. Teknologi väljs efter den kliniska applikationens krav.

- Variantkvalitet [Hela processen]

Variantbestämningens kvalitet beror primärt av enskilda basers kvalitet och alignmentkvalitet. Även andra problem än de som fångas av dessa kan dock vägas in, såsom strängskevhet för till varianten bidragande läsningar. Genom att studera variantkvalitetsparametrar för ett stort antal varianter i olika prover kan en empirisk felsannolikhet utarbetas (t ex som GATK VSQR).

- Variantantal [Hela processen]

Antal varianter jämfört med referens i en målregion kan användas som en indikator, även om den är individvariabel. Mycket få varianter respektive extremt många kan tyda på problem med sekvensering, biblioteksberedning, inpassning eller variantbestämning.

- Andel heterozygoter/homozygoter [Hela processen]

Andel homozygoter/heterozygoter följs med fördel. Dessa kan t ex indikera problem med provblandning (många heterozygota varianter, få homozygota) eller brist på unikt templat (många homozygota varianter).

- Transition/transversionskvot [Hela processen]

Transition/transversionskvoten tenderar att hålla sig relativt konstant för ett målområde. Avvikelser kan tyda på kemiska problem eller större problem på datorsidan.

Sammanfattande rekommendation

Sammanfattningsvis bör inte avsteg göras från de tidigare rekommendationer från riktlinjer för sekvensering som SFMG upprättat. De kan utan allt för långtgående modifiering appliceras även vid storskalig sekvensering, liksom god sed i validering och verifiering som tillämpas vid ackrediterade kliniskt genetiska laboratorier.

Flertalet kliniker har idag utfört valideringar och verifieringar av storskaliga tekniker inom ramen för sin Swedac-ackreditering. Vi uppmuntrar den som önskar se sådana exempel att kontakta t ex Klinisk genetik i Göteborg, Linköping eller Uppsala för att få ta del av aktuella sådana.