Helgenomsekvensering (WGS) – fördjupad teknisk information

Den aktuella analysen innebär sekvensering av hela det mänskliga genomet (helgenomsekvensering, WGS), och har mycket god teknisk förmåga att detektera såväl små genetiska varianter (SNVs och mindre indels) som större kopietalsavvikelser (CNVs, deletioner och duplikationer). Avseende varianter i mitokondriellt DNA är metoden validerad för detektion ner till 5% heteroplasmi (variantallelfrekvens). Då kunskapen om icke-kodande genetisk variation är begränsad så kommer klinisk tolkning för s.k. singelprover, d.v.s. när endast symptomatisk patient finns tillgänglig, främst vara begränsad till kodande sekvens inom angiven genlista (se individuellt labsvar), och redan välbeskrivna patogena varianter i icke-kodande sekvens. Vid utredning där det bereds tillgång till föräldraprover (s.k. trio) utförs en bredare analys, vilken inkluderar samtliga kodande gener. Se nedan för en mer detaljerad beskrivning av det bioinformatiska flödet vid WGS.

Varianter vilka tolkas som av oklar klinisk signifikans rapporteras ej, såtillvida det inte motiveras av möjlighet till uppföljande klinisk analys eller segregationsanalys. Framtida kunskap kan komma att modifiera aktuell tolkning, varför det är möjligt att med förnyad remiss begära reanalys av framtagen data.

Metoden detekterar även trinukleotidexpansioner (t.ex. Fragilt-X, Dystrofia myotonika, Huntington, C9ORF72-relaterad ALS/FTD, samt flera spinocerebellära ataxier) men då denna del av analysen ännu ej är validerad (ingår ej i ackrediteringen) så verifieras alltid positiva fynd med separat metod. Analysen innefattar för tillfället inte komplexa strukturella varianter.

Syftet med WGS-analysen är att upptäcka orsaken till ett tillstånd hos patienten, inte att upptäcka bärarstatus av recessiva sjukdomar. Bärarstatus av en given sjukdom kan förfrågas i specifika situationer, t.ex. vid konsanguinitet, eller om partner till patienten har en molekylärgenetiskt bekräftad sjukdom/bärarskap.

Nomenklatur i labsvar följer HGVS nomenklatur guidelines (http://varnomen.hgvs.org) och variantklassificering följer huvudsakligen riktlinjer från ACMG (PMID: 25741868, PMID: 31690835, PMID: 32906214). I de fall det finns genspecifika klassningskriterier används dessa.

Bioinformatiskt flöde vid WGS

Preparering av sekvensbibliotek för helgenomsekvensering (WGS) görs med TruSeq DNA PCR-Free (Illumina) och sekvensering utförs med NovaSeq 6000 (Illumina). Från sekvenseringsinstrumentets råa basbestämning (.bcl, binary basecall) genereras sekvensdata i fastq-format med hjälp av verktyget bcl2fastq. Sekvensdata inpassas mot det humana genomet (GRCh38 som referenssekvens) med mjukvaran BWA (Burrows-Wheeler Aligner).

Duplicerade läsningar, definierade som läsningar med ursprung från samma DNA-molekyl, identifieras med Sentieon LocusCollector/Dedup och undantas från nedströms analys. Baskvalitetsvärden rekalibreras med hjälp av Sentieon QualCal. Den slutgiltiga sekvensinpassningsfilen (i bam-format, Binary sequence alignment map) används för att beräkna kvalitetsparametrar med hjälp av Sentieons QC-moduler.

Avseende SNV- och Indel- varianter bedöms endast varianter lokaliserade i eller inom 20 baspar från exon (enligt senaste Ensembl-version) samt introniska/intrageniska varianter eller varianter i icke-protein kodande gener tidigare rapporterade som patogena eller sannolikt patogena i databasen ClinVar (https://www.ncbi.nlm.nih.gov/clinvar). Eftersökning av kliniskt relevanta kopietalsvarianter (CNVs) och genotypning för detektion av t.ex. uniparentell disomi (UPD) utförs däremot för hela genomet. Manuell variantbedömning görs primärt i Scout (https://clinical-genomics.github.io/scout).

Analysflöde för variantbestämning samt annotering gällande SNV och små indels

Nukleärt DNA
Variantbestämning utförs med hjälp av verktyget Sentieon DNAscope enligt rekommenderat arbetssätt, (https://support.sentieon.com/manual/DNAscope_usage/dnascope). Detekterade varianter annoteras och rankas i ett annoteringsflöde bestående av verktygen VEP[1], vcfanno[2], CADD[3] och Genmod (https://github.com/moonso/genmod). Genmod är en flerstegsprocess där all information som krävs för att beräkna rankscore (t.ex. populationsfrekvenser, proteinpåverkan, och för trioanalys - nedärvningsmönster) tillfogas VCF-filen för att i sista steget slås samman i en rankscore. Slutligen laddas den annoterade variantlistan in i Scout för vidare tolkning. Data används även för kontroll att samstämmighet för trio och kön stämmer mellan genotypdata och den angivna informationen. För singelanalys görs endast en kontroll av könsinformation.

Mitokondriellt DNA
Variantbestämning utförs med hjälp av verktyget Mutect2, i mitochondria mode, som är en del av GATK best practice mjukvarupaket (https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2). Varianter annoteras först separat från ovanstående nukleära flöde, med hmtnote, och följer därefter samma steg som annotering av nukleära varianter med VEP, vcfanno, CADD och Genmod. Mitokondriella varianter undantas från nedärvningsmönster, compounds samt splitspåverkan. För mitokondriella varianter används gnomAD 3.1 som populationsfrekvensannotering. Haplotypering görs med haplogrep, och presenteras som bilder.

Analysflöde för variantbestämning samt annotering gällande strukturella varianter

Nukleärt DNA
Variantbestämning utförs med hjälp av tre olika variantbestämmare; GATK[4], TIDDIT[5] och Manta[6]. GATK använder sig av läsdjupet över kromosomer för variantbestämning, specialiserat för att hitta stora strukturella varianter. GATK gör felkorrigering och bias-korrigering med hjälp av en normalpanel, vilket minskar antalet falska positiva avrop. TIDDIT och Manta använder sig av diskordanta parade läsningar i kombination med split-read-

läsningar för variantbestämning, och har därför bättre känslighet vid små avvikelser samt möjliggör detektion av kopieneutrala strukturella varianter. Genom att kombinera flera olika verktyg (algoritmer för variantbestämning) för att detektera strukturella avvikelser uppnås en högre tillförlitlighet och risken att missa detektion av kliniskt relevanta CNVs reduceras.

Efter variantbestämning sammanfogas alla varianter som överlappar med 70% från de tre olika variantbestämmarna. Detta sker med mjukvaran SVDB (lokal): de varianter vilka överlappar tidigare bestämda varianter med 70% sammanfogas till en variant. Det utförs även en annotering från SVdb, en lokal databas för strukturella varianter från helgenomsekvens.

Detekterade varianter annoteras och rankas i ett annoteringsflöde bestående av verktygen VEP[1], AnnotSV[7], Pre-score (lokalt Pearlscript), Genmod (se beskrivning ovan under SNV/Indels) och Compound finder (lokalt Pearlscript). AnnotSV annoterar strukturella varianter med bland annat dbvar, gnomAD-SV och gener i databas OMIM-Morbid Genes.

Varianttolkning och manuell annotering utförs i Scout med stöd av Alamut Visual (https://www.interactive-biosoftware.com/alamut-visual) och IGV (https://software.broadinstitute.org/software/igv) samt lokalt utvecklade visualiseringsverktyg (bl.a. WGS-gSNP, WGS-POD och Gens).

Mitokondriellt DNA
Variantbestämning utförs på samma sätt som för nukleärt DNA, med skillnaden att GATK inte används. Utöver Manta och TIDDIT används även en fristående variantbestämmare, eKLIPse, som genererar cirkulära plottar för detektion av mtDNA rearrangemang (kopietalsförändringar). Verktyget eKLIPse ger brytpunkter för avvikelser liksom deras heteroplasminivå.

Referenser

1. Ensembl Variant Effect Predictor (VEP): https://www.ensembl.org/info/docs/tools/vep/index.html
2. vcfanno https://github.com/brentp/vcfanno
3. Combined Annotation Dependent Depletion (CADD): https://cadd.gs.washington.edu
4. https://gatk.broadinstitute.org/hc/en-us/articles/360035531152--How-to-Call-common-and-rare-germline-copy-number-variants
5. Eisfeldt J. et al., TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Research*, 2017. (PMID: 28781756)
6. Chen X. et al., Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 2016. (PMID: 26647377)
7. Geoffroy V et al., AnnotSV: an integrated tool for structural variations annotation. Bioinformatics, 2018. (PMID: 29669011): https://lbgi.fr/AnnotSV