Alissa DSS replacement inhouse pipeline requirements

## panHO requirements

1. QC requirements:
   1. Essential:
      1. QC html report is uploaded to Alissa for sharing sample specific coverage data with Wessex.
   2. Nice to have:
      1. Look at coverage represented by gene rather than just exon (by machine rather than exon/gene?). This was available in the MiSeq pipeline but is not available with Alissa. JL/NB previously tried to develop this? Further information required as to history.
2. Annotation requirements:
   1. Essential:
      1. RefSeq annotation – **must be kept updated to latest version.**
      2. GnomAd population data to filter polymorphisms – **must be kept updated to latest version.**
         1. There are caveats for certain variants where you would need to see the population frequency data.
      3. Revel score **– would help a lot.**
   2. Nice to have:
      1. ClinVar annotation.
      2. COSMIC annotation is used by Wessex only.
   3. Future desirable features:
      1. Annotation with disease/gene specific databases e.g. TP53 database. Clinical Scientists can provide a list of these disease/gene specific databases.
3. Filtering settings:
   1. EA to extract filtering settings from Alissa decision tree for each pipeline/panel; JB/NF to review list for each panel once completed.
   2. Once extracted confirm these with clinical scientist.
      1. **Scientists’ responsibility to clearly define/check the filtering settings (e.g. %VAF).**
      2. BI to implement filtering decision tree in excels.
4. Output:
   1. Essential:
      1. Provide filtered excel report (reduces number of variants and wastes less time during variant interpretation)
      2. Include: Chr & co-ordinates; Ref and Alt allele, gene, Exon, Transcripts, HGVS nomenclature
   2. Nice to have:
      1. Include sample ID for each row of data. This makes it easier to compare two patients if necessary because each row has a sample ID. (This should be easy to implement).
      2. Include the TSV file created from the ‘.cnv.gff3.tsv’ by JL/NB as a tab within each excel report. (If this TSV file already contains the data required, then it should be fairly easy to add this to a tab within an excel file).

General Points:

* Start with SNPs as these are more straight forward.
* CNVS and translocations will require further input from the clinical scientists as these are more complex – may require tweaking of the Dragen pipeline.
* Look at an example of an old TSMP report.
* Look at example of Alissa PAWS excel report.
* TSO500 may require reanalysis with different panel – ask TSO500 team about this.

Currently:

* panHO SOP – zipped file uploaded to Alissa:
  + concatenated VCF
  + HTML QC report – sample specific coverage metrics
  + ‘.blf’ file – bam file location
  + ‘.cnv.gff3.tsv’ file containing CNV call details
  + Manifest Text file for Alissa Interpret to recognise uploaded files.

Variant annotation tools :

[JCM | Free Full-Text | Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics (mdpi.com)](https://www.mdpi.com/2077-0383/9/1/132)

Nirvana (Illumina annotation engine on DRAGEN

Anovar

Others see 🡺 [JCM | Free Full-Text | Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics (mdpi.com)](https://www.mdpi.com/2077-0383/9/1/132)

[Nirvana | Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics](https://dl.acm.org/doi/10.1145/3107411.3108204)