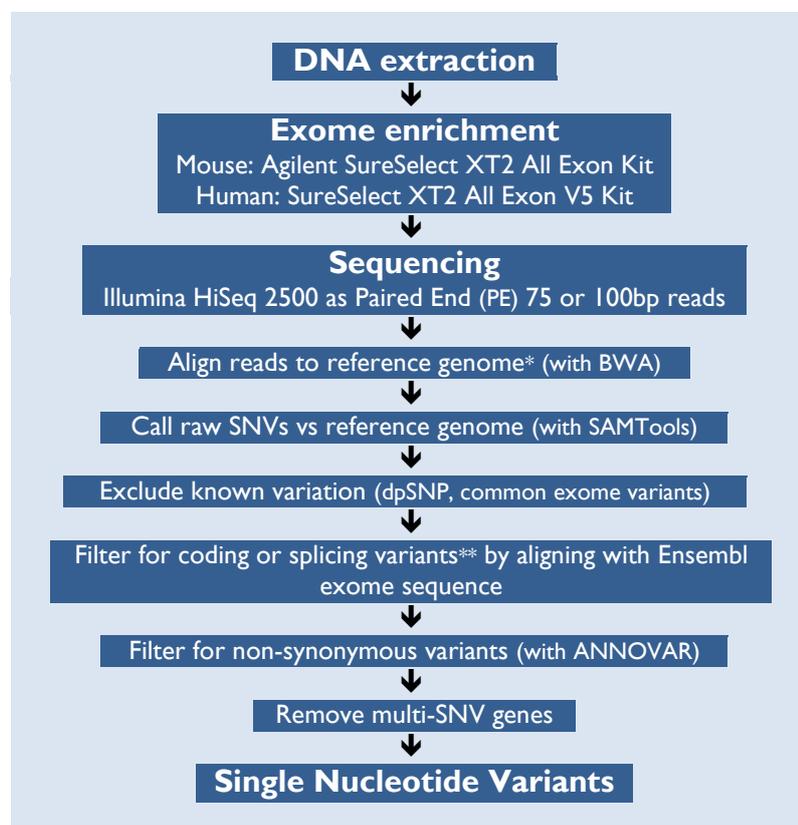


APF Exome Sequencing Service Overview

Exome Sequencing Pipeline



The bioinformatics workflow brings together a number of open-source analysis tools and employs a Perl code-base to perform custom filtering, reporting and job process control. The APF bioinformatics pipeline for Single Nucleotide Variant (SNV) and indel identification utilises the ENSEMBL exome sequence for human and mouse to map the data.

*reference mouse genome (mm10/GRCm38)

** splicing variants refer to changes that lay in potential splice donor-acceptor sites immediately adjacent to exon boundaries (out to 10 intronic bases)

Sequencing Analysis Output

The sequencing output is six files for each sequenced exome.

readReport_summary	Gives the number of reads obtained and the percentage of reads that mapped to the target Ensembl exome.
snpList_summary	The actual filtered SNV list. Annotation with SNV allele frequency, CCDS gene id, Ensembl gene id, Online Mendelian Inheritance in Man (OMIM) loci, mouse mutant phenotype, expression patterns and PolyPhen2 scores for all novel SNVs
snpReport_summary	Lists the filters used in the analysis, the number and percentage of SNVs that passed each filter, the final number of passed SNVs and any genes with more than one variant (SNVs in these genes could be a result of mis-mapping due to high homology with other genes).
exonReport_summary	Provides a list of exons/regions that are covered by less than three reads.
indelList_summary	The filtered indel list.
indelReport_summary	Lists the filters used in the analysis, the number and percentage of indels that passed each filter, the final number of passed indels and any genes with more than one variant (indels in these genes could be a result of mis-mapping due to high homology with other genes).

Variant validation

The APF also offer a SNV validation service to customers. SNVs are validated using Amplifluor assays (Chemicon, Temecula, CA). Primers are designed using a semi-automated pipeline and results are provided in excel format. If an amplifluor assay does not work the researcher has the option of requesting that a Sanger sequencing assay be designed to the SNV.

Data Retention Policy

Sequencing data	Due to the large size and high volume of sequencing data, the APF has very limited capacity for data retention.
Intermediate analysis files	The intermediate files will not be maintained at the APF. If users would like a copy of the intermediate analysis files, they must request this when submitting samples AND provide a portable hard drive for the file transfer. There will be a cost associated with the data transfer.
Analyzed Sequence files (fastQ & aligned data)	This data will be stored at the APF for 3 months. For operational purposes, the APF may retain data beyond six months, but cannot guarantee retrievability after six months.
Reports	These will be provided to the researcher following final QC. The files will also be stored at the APF for 5 years.
QC data	All Agilent Bioanalyser files will be stored for 5 years at the APF and can be provided upon request.