SERVICE GUIDE

Whole Exome Sequencing

CustomerCare@agrf.org.au
1300 247 301
www.agrf.org.au

SVG2209WES

Service Guide:
Whole Exome Sequencing

Service Guide:
Whole Exome Sequencing

1.0 Overview

Whole Exome Sequencing (WES) is a method used to selectively sequence the exons or protein coding region of the genome. While exons constitute approximately 1% of the human genome, an estimated 85% of disease-related variants occur within the protein coding region of the genome. Whole exome sequencing provides high coverage and high confidence in small germline variant calls, including single nucleotide variations and small insertion/deletions for the study of rare mendelian disorders, complex disease, and cancer.

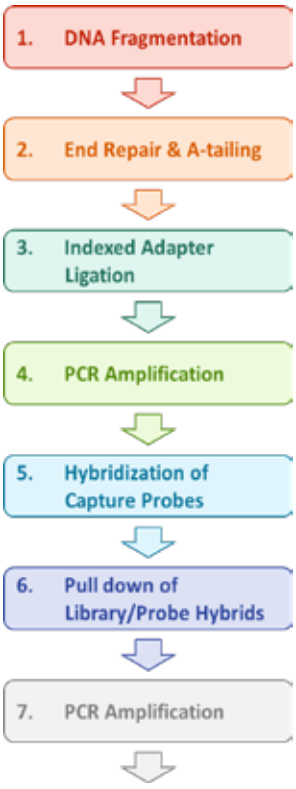
2.0 Sequence Capture

Whole exome sequencing utilises a process of hybrid capture (or sequence capture) to enrich specific targets of the genome. In this process, a pool of biotinylated oligonucleotide probes is selectively hybridised to a targeted exon of a fragmented DNA sample. DNA fragments bound to probes are then pulled down using streptavidin magnetic beads. After the removal of un-bound DNA, the captured DNA fragments are processed for sequencing.

At AGRF, we use the Twist Exome V2 panel with Twist EF library preparation workflow for our new generation, low-cost research exome workflow. This workflow requires a minimum of 50 ng input with high quality DNA (integrity >7). If you are working with degraded material or FFPE please let us know as we instead use the IDT xGen cfDNA + FFPE library prep workflow as the front-end library preparation with the Twist Exome V2 panel.

We also use the SureSelectXT Low Input target enrichment system from Agilent for our Clinical whole exome sequencing service. The SureSelectXT Low Input protocol provides flexible options for processing high quality DNA and low quality FFPE DNA over a range of inputs (10ng to 200ng).

Figure 1: Summarised workflow for Exome Sequence Capture.



3.0 Germline Variant Calling

AGRF provides a comprehensive service for WES, including bioinformatics. We use the Illumina DRAGEN™ algorithm for alignment, duplicate marking and small variant calling (SNV/INDELS). Germline variant calling utilises the following steps: Identification of active region, localised *de novo* assembly of haplotypes, read likelihood estimations and variant genotyping based on a diploid model. These algorithms result in very high accuracy for both Single Nucleotide Variants (SNV) and small Insertion and Deletions (Indels). Illumina's DRAGEN™ also provides variant quality scores and post-VCF variant filter annotations. The reconfigurable Bio-IT processor of DRAGEN™ also substantially improves the speed of analysis. For more information on DRAGEN™, please visit this webpage.

Figure 2: Summarised workflow for Exome variant calling.

4.0 AGRF Services

AGRF provides three comprehensive packages for human WES: Standard, Deep, and Clinical. AGRF can also provide tailored options for specific sequencing depths and other biological systems (subject to availability and minimum batch sizes).

Service	Description
Twist Comprehensive v2 Human WES – Standard Research	• Twist Comprehensive Exome V2 (36 Mbp Target) • Illumina Sequencing – 150bp paired end reads • Min 5 Gbp (mean on-target coverage 40-80X)* • Human genome alignment and germline variant calling
Twist Comprehensive v2 Human WES – Deep Research	• Twist Comprehensive Exome V2 (36 Mbp Target) • Illumina Sequencing – 150bp paired end reads • Min 10Gbp (mean on-target coverage 80-160X)* • Human genome alignment and germline variant calling
Agilent CREv2 Human WES – Deep Clinical	• ISO15189(2012) Accredited • Agilent SureSelectXT Low Input Clinical Research Exomev2 (68Mbp design) • Illumina Sequencing – 150bp paired end reads • Min 12Gbp (mean on-target coverage >100X*) • Human genome alignment and germline variant calling
Agilent CREv2 Human WES – Ultra Deep Clinical	• ISO15189(2012) Accredited • Agilent SureSelectXT Low Input Clinical Research Exomev2 (68Mbp design) • Illumina Sequencing – 150bp paired end reads • Min 24Gbp (mean on-target coverage >200X*) • Human genome alignment and germline variant calling

Table 1: Summary of standard human whole exome sequencing packages available at AGRF.

*Mean on-target coverage is highly dependent on sample quality, integrity, and input. The values specified should be used as a guide only and assumes high quality (integrity >9) material with a 100ng input.

Output	Description
FASTQ	Data – raw sequencing output, Illumina FastQ format
BAM	Data – indexed BAM alignment
Hard Filtering VCF	Data – indexed variant calls – includes internal variant filter status, e.g. PASS
VEP	Data – annotation and html summary
Post Analysis QC report	QC Metric – Coverage, yield etc.

Table 2: Summary of deliverables for human whole exome sequencing.

5.0 Data Outputs

Clinical whole exome sequencing also includes additional Quality Control, including gender and contamination checks, as well as individual gene coverage metrics. For clinical whole exome sequencing, the bioinformatic analysis uses the Burrows-Wheeler Aligner Tool (BWA mem) for clinical germline variant calling, the bioinformatic analysis workflow also uses the Illumina's DRAGEN TM suite for alignment and variant calling.

6.0 Sample Requirements

High quality DNA is required for optimal results for exome analysis. Please refer to the following for our recommendations for sample quantity and quality.

6.1 Sample Quantity

AGRF requests 500ng of high-quality DNA per sample (20µL per sample and >20ng/µL in nuclease-free water or TrisHCl/Qiagen EB).

If you are unable to supply this amount of material or if you are working with low-input OR low quality (including low-quality FFPE or cfDNA) samples, we request 100 ng of DNA per sample (20µL per sample and >2.5ng/µL in nuclease-free water or TrisHCl/Qiagen EB).

If obtaining high quality or sufficient material is difficult, please contact us to discuss options as we may still be able to process your samples using protocol modifications.

AGRF recommends the use of fluorometric quantitation methods (e.g. Qubit) for the most accurate quantitation.

These requirements apply to both the Twist and SureSelect workflows.

6.2 Sample Quality

We will perform quality control prior to commencing a project, however, we recommend you check the DNA before submission.

6.3 Sample Purity

Purity of nucleic acid samples can be assessed by measuring the absorbance spectra on a spectrophotometer (e.g. NanoDrop). The ratio of absorbance values 260nm/280nm or 260nm/230nm provides an estimate of sample purity or the presence of common contaminants. Purified DNA is expected to have a ratio at A260/280 of ~1.8.

Ratio	Target	High/Low Ratio Indications
A260/280	1.8	<1.6 Possible protein contamination, or very low conc. of DNA (<10ng/µl) >1.8 Possible RNA contamination
A260/A230	>1.6	Residual guanidine from the extraction protocol Carryover of carbohydrates (e.g. plant polysaccharides)

Table 3: Recommendations for sample purity as assessed by absorbance spectra.

6.4 DNA Integrity

DNA Integrity can be assessed by using either agarose gel electrophoresis or a capillary electrophoresis system such as an Agilent TapeStation or Bioanalyzer or equivalent. DNA should appear as a clear, high molecular weight band at the top of the gel. There should be no indication of RNA contamination (as evident

by faint dual banding towards the bottom of the gel).

7.0 Sample Returns/Discards

Samples are stored with AGRF for 3 months after you receive your data except for our clinical WGS and clinical exome services which are stored for 1 year. If you wish for your samples to be returned, you must discuss this with your account manager during quoting or contact us after you receive your data. At the completion of your project, we can either:
1. Return your samples by courier at ambient (please ask your account manager for a quote).
2. Return samples by courier with dry ice (please ask your account manager for a quote).

If we are not notified within the specified time frame, samples will be automatically discarded.

8.0 How to Submit Samples

Online Submission

- In the client portal, select 'Next Generation Sequencing' from the service dropdown menu.
- Enter your species (Human only) and submission format (tube or plate).
- Complete and upload the template file.
- Submit the form and print the submission receipt to be included with your sample package.

Packaging of Samples

- DNA samples can be shipped at room temperature via courier or express post.
- To prevent leakage in transit; please use parafilm to seal your tubes. Plates should be heat sealed or sealed with strip caps.

Send your samples to one of the addresses below:

Physical address (courier):
ATTN Next Generation Sequencing
AGRF
Level 13, Victorian Comprehensive Cancer Centre
305 Grattan Street
Melbourne, VIC 3000

Postal address (mail):
ATTN Next Generation Sequencing
AGRF
VCCC Loading Dock
14 Flemington Road
North Melbourne, VIC 3051

9.0 Quality Statement

The Clinical workflow for our WES service follows strict requirements of ISO15189: 2013. AGRF Ltd is accredited as a Medical Testing laboratory according to the ISO15189: 2013 standard by the National Association of Testing Authorities (NATA). Compliance is monitored at regular reviews and internal audits.

All other works carried out by AGRF are performed following the strict requirements of ISO17025. AGRF Ltd is accredited in the field of Biological Testing (Scope: DNA Analysis) according to the ISO17025 standard by the National Association of Testing Authorities (NATA). Staff follow Standard Operating Procedures, which define their responsibilities and provide guidance on achieving standards; compliance is monitored at regular reviews and internal audits. All work is supervised by a person with relevant qualifications and is checked while in progress and upon completion to ensure that it met the necessary ISO17025 standards.

AGRF is a Certified Service Provider (CSP) for Agilent capture based targeted enrichment and Illumina sequencing. CSP programs are collaborative service partnerships between facilities and vendors to ensure best practice and the highest data quality.