Frequently Asked Questions

Whole Exome Sequencing and Analysis

Q1. What is Whole Exome Sequencing ?

A1. Whole Exome Sequencing (WES) is an efficient strategy to selectively sequence the coding regions (exons) of a genome, typically human, to discover rare or common variants associated with a disorder or phenotype [1, 2]. By focusing sequence production on exons, which represents ~2.5% of the human genome, many more individuals can be examined at significantly reduced cost and time compared to sequencing their entire genomes. The most common methods rely on hybridization by oligonucleotide probes to 'capture' targeted DNA fragments, thereby enriching for exonic sequences. Targeted exonic sequences include well-established annotated coding and non-coding exons. Regions not within close proximity, on the order of 100-bases, of the targeted regions are not sequenced. Therefore, variants within introns, promoters or inter-genic regions are generally not detected.
Note, DNA samples derived from living humans must be consented for WES before acceptance at NISC for sequencing.

Q2. How is WES performed at NISC ?

A2. NISC employs a solution-based probe hybridization protocol to capture (enrich for) exonic sequences from the DNA sample. The current whole-exome capture kit used at NISC is the IDT xGen Exome Research Panel which targets a total of 39 Mb. In brief, the DNA is sheared to a uniform size appropriate for sequencing, fragments are captured by probe hybridization, and then amplified before sequencing on an Illumina NovaSeq 6000 instrument. NISC continually evaluates improvements in these technologies, and implements those that represent reduction in cost and time or increase performance.

Q3. What material should I send for WES ?

A3. We need a minimum of 150 ng of highly-purified genomic DNA (0.5 µg preferred) in a volume of 50 µl or less for WES. Samples should be submitted in 1.5-1.7 ml microfuge tubes (example: VWR cat. no.89000-028) or 2 ml screw cap tubes (example: Sarstedt cat. no. 72.694.007). Please DO NOT send samples in 0.5 or 0.2 ml tubes. To ensure that each sample is uniformly pure and free of infectious agents, we strongly recommend that all DNAs be phenol:chloroform extracted before submission. A simple protocol is available from NISC. Ref:
www.nisc.nih.gov/docs/gDNA_submission_exome_cc.pdf

Frequently Asked Questions

Q4. How should the DNA be qualified ?

A4. The investigator must submit an image of an analytical agarose gel or a trace as evidence the DNA is of good integrity and the appropriate molecular weight for the sequencing approach. We highly recommend Qubit for quantitation of the DNA sample, since it uses a double-strand DNA-specific method. UV absorption methods, e.g., using a NanoDrop spectrophotometer, can drastically overestimate the concentration of DNA due to RNA and small molecule contamination.

Q5. Can DNA extracted from FFPE tissue be used in WES ?

A5. DNA from FFPE is always damaged to some degree by the harsh chemical treatments used to preserve the tissue. The degree of damage is influenced by a number of factors, the protocol used in fixing the tissue, the length of time the tissue was stored, and the DNA extraction protocol used. Some damage can be repaired, but not all. For instance, abasic positions cannot be restored. As with all DNA samples, FFPE-derived DNA should be assessed for integrity by agarose gel electrophoresis or a trace. This analysis is a strong predictor of how useful the WES data will be. The greater the degradation, the poorer the results. In general, you can expect somewhat poorer data from FFPE-derived DNA.

Q6. How long are the reads for WES analyses ?

A6. Typically, NISC generates read lengths of 150 bases on a NovaSeq 6000. Paired-end reads generate a total of 300 bases of sequence (150b from each end) from each fragment in the library.

Q7. How many reads are required for WES analyses ?

A7. Currently, we target a minimum of 25 million paired-end 150 base reads which will yield an average read-depth of at least 75x.

Q8. How are variants called in WES analyses ?

A8. Sequence reads produced for a sample are aligned to the human reference sequence and the results stored in BAM format. A custom analysis program, MPG (Most Probable Genotype), processes this information using a probabilistic Bayesian algorithm, calling genotypes at all reference positions at which there are high quality bases from the aligned sequence reads [4]. The likelihood of each possible genotype from the observed sequence data is calculated and given an MPG score, where MPG 10 is considered accurate. These genotype calls have been compared against Illumina Human 1M-Quad genotype chips,

and genotypes with a MPG score of 10 or greater show >99.89% concordance with SNP chip data [4].

Q9. What data are returned by NISC ?

A9. All variants, genotypes, and annotations are delivered to the investigator in tab-delimited format compatible with VarSifter [5], a java-based genotype viewer, available from NISC. The file can also be imported to Excel. The VarSifter file contains all discovered variants with genotypes of all samples sequenced, as well as gene locations (5' UTR, 3' UTR, coding−synonymous, nonsynonymous, or stop, splice site, or intron).

References:

1. Biesecker L (2010) "Exome sequencing makes medical genomics a reality." *Nature Gen.* 42, 13-14.
2. Illumina (2013) "An Introduction to Next-Generation Sequencing Technology." www.illumina.com/documents/products/Illumina_Sequencing_Introduction.pdf
3. Teer, JK *et al.* (2010) "Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing." *Genome Res.* **20**: 1420-1431
4. Sims, D., *et al.* (2014) "Sequencing depth and coverage: key considerations in genomic analyses." *Nature Rev. Genetics* **15**: 121-132.
5. Teer, JK, *et al* (2012) "VarSifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer." *Bioinformatics* **28**: 599-600.