
Frequently Asked Questions

Custom Capture Sequencing

Q1. What is custom capture sequencing ?

A1. Custom Capture Sequencing (CCS) is an efficient strategy to selectively sequence specific (targeted) regions of interest in the genome to discover rare or common variants associated with a disorder or phenotype [1, 2]. By focusing sequence production on targeted regions of the genome, many more individuals can be examined at significantly reduced cost and time compared to sequencing their entire genomes. Custom capture sequencing gives the option of targeting exons in selected genes or tiling across an entire region of the genome. The cumulative size of the targeted regions of interest typically ranges from ~500 kb – 50 Mb.

Q2. How is CCS performed at NISC ?

A2. The standard starting point for a custom capture project is the identification of a region or regions of interest. The NISC staff then uses this information to develop a custom capture design which typically consists of oligonucleotide probes for hybridization-based capture or multiplex PCR primers that can be used to enrich for the targeted region(s) of the genome. NISC provides the investigator for review a summary of the design efficiency, i.e., predicted coverage of the target region(s). Based on feedback, the design is modified as necessary. Once the investigator has given final approval of the design, then the custom capture kit is ordered. Note that the time from beginning the design process to receipt of a custom capture kit at NISC is on the order of 10 weeks. Only then can NISC start sequencing samples previously received. Custom capture projects typically include 24-576 samples.

Q3. What material should I send for CCS ?

A3. We need a minimum of 1.5 µg of highly-purified genomic DNA (5 µg preferred) in a volume of 120 µl or less. 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 require 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 as evidence the DNA is of good integrity, i.e., not just a 'fuzzy blob' of low molecular weight. 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. How long are the reads for CCS analyses ?

A5. Typically, NISC generates read lengths of 125 bases on a HiSeq. Paired-end reads generate a total of 250 bases of sequence (125b from each end) from each fragment in the library.

Q6. How many reads are required for CCS analyses ?

A6. We target a number of reads calculated to produce an average read-depth of greater than 100X coverage [3].

Q7. How are variants called ?

A7. 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].

Q8. What data are returned by NISC ?

A8. 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).

Frequently Asked Questions

References:

1. NimbleGen (2013) “SeqCap EZ Library.”
www.nimblegen.com/products/lit/05227887001_SeqCapBroch_Jan2013.pdf
2. Illumina (2013) “An Introduction to Next-Generation Sequencing Technology.”
www.illumina.com/documents/products/Illumina_Sequencing_Introduction.pdf
3. Sims, D., *et al.* (2014) “Sequencing depth and coverage: key considerations in genomic analyses.” *Nature Rev. Genetics* **15**: 121-132.
4. Teer, JK *et al.* (2010) “Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing.” *Genome Res.* **20**: 1420-1431.
5. Teer, JK, *et al* (2012) “VarSifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer”. *Bioinformatics* **28**: 599-600.