Target Region Sequencing FAQ

2011-04-28 (Version 1)

Contents

1. What is target region sequencing, and how do you initiate a target region sequencing project with BGI?

   Target region capture is the enrichment of specific regions (e.g., MHC region) or specific genes by microarray hybridization using NimbleGen Sequence Capture array or solution hybridization using the Agilent Sure-Select™ system based on designed probes which are designed according to the base sequence of genomic regions of interest. You only need to provide genomic DNA samples and a candidate genes list. BGI provides a complete solution including probe customization, hybridization, high throughput sequencing, and bioinformatics analysis.

   Information such as SNPs, InDels can be obtained after target region sequencing.

2. What are the advantages of target region sequencing?

   A target region is comparatively smaller than the whole exome, and as such partners can spend less and get more information for the region of interest. In addition, target region sequencing can also focus on non-coding regions that may be related to diseases and is extremely suitable for case/control studies with large sample sizes. Target region sequencing is more flexible and has extensive applications.

3. Are there any specific study proposals for target region sequencing?

   Theoretically, the more samples, the higher the chance of discovering susceptible genes or variants. These are some recommendations for different research goals:

   If you want to search for new variants, a large sample size and regular sequencing depth is recommended.

   If you intend to validate the candidate loci, fewer samples and deeper sequencing is recommended.

4. What are the sample requirements?

   The sample requirements are as follows:

   Sample type must be DNA without degradation and RNA contamination

   Sample quantity should be ≥ 30μg if using NimbleGen Sequence Capture Array and the target region is larger than 17Mb or ≥ 6μg if using Agilent Sure-Select system or NimbleGen Sequence Capture Array and the target region is smaller than 17Mb.

   Sample concentration should be ≥ 50ng/μl.

   Sample purity should have an OD260/280 = 1.8 - 2.0.

5. Can the samples be pooled without indexing to reduce the cost?

   Generally, we attach an index to isolate samples from the mixtures of several samples. Sample pooling without indexing is not suggested, as this will result in biased sequencing and greatly increase the false positive and false negative ratio of the variants detection. In addition, the non-uniformity of mixed samples may also result in an unsuccessful analysis.

6. Are there any cases in target region sequencing in which probes design can be exempt for target region sequencing?

   No matter where the target region is located (even if there is overlap with the exon regions), customized probe design is necessary. Detailed sequence information of the target region must be provided to determine if it is suitable for chip design. In addition, provide the target sequence information using the hg19 version, which will be the reference for subsequent bioinformatics analysis.

7. What bioinformatics analysis results are generated with target region sequencing?

Analysis Types	**Analysis Techniques Available Based on Project Requirements**		**Notes**
Standard Bioinformatics Analysis	1) Data filtering (removing adaptors contamination and low-quality reads from raw reads) 2) Alignment, summary of data production (based on 1) 3) SNP calling, annotation and statistics (based on 1 and 2)		reference version is Hg19
Advanced Bioinformatics Analysis	Variation detection	4) InDels calling, annotation and statistics (BWA software will be applied for alignment if chosen)	based on 1 through 3 of standard bioinformatics analysis
	Non-disease	Population and Evolution: 5) Population SNP calling 6) Population InDel calling, based on analysis 4) 7) Haploview: a) Linkage disequilibrium b) Haplotype prediction 8) Positive selection signals detection	based on 1 through 3 of standard bioinformatics analysis

	Disease-related	*Cancer* 9) Somatic mutation (SNP / InDels) screen, based on analyses 3 and) (tumor-normal paired samples required) 10) SNV detection for tumor-normal paired samples; 11) Filtering with the known databases like cosmic, dbSNP etc., based on analyses 9 and 10 12) Amino acid substitution prediction (Sift, Polyphon-2) based on analysis 9 13) GO enrichment analysis for selected genes, pathway enrichment analysis based on analyses 9) and 10) 14) Family-based analysis for pedigree samples *Complex Disease* NGS-GWAS: 15) Population SNP calling and minor allele frequency (MAF) estimation 16) PLINK-based analysis (for large scale samples or reasonably selected samples of fewer number) 17) Genotype-based analysis (requires relatively high sample number) 18) Gene-based associated analysis 19) Haploview-based analysis (cooperation preferred for R&D) *Others* 20) Positive selection signals detection 21) Genetic analysis of familial aggregation disease based on sequencing and linkage analysis (cooperation preferred)	based on 1 through 3 of standard bioinformatics analysis
Personalized Analysis	Details can be determined based on discussions between the customer and BGI		based on 1 through 3 of standard bioinformatics analysis

8. What are the requirements for the fragment length of the PCR product?

   There are different processing methods and recommendations for different lengths of fragments. Note the following regarding fragment length:

If the fragment is < 100bp, contamination of the adapter increases during sequencing and leads to wasted data. Also there is a chance of losing fragments during the addition of adapters if the fragment size is too small.

If the fragment is from 100 to 200bp, the adapters can be added directly and sequenced, and there is no need for reconstructing library by fragmentation. Sequencing can be done using PE91 or PE101.

If the fragment is from 200 to 1000bp, construction of the library is not recommended. It is possible to sequence the two ends of the PCR products. Sequencing can be done if you ligate the PCR fragments with the reconstructed length > 1Kb .However, this method creates bias, and the goals of sequencing cannot be achieved.

If the fragment is > 1Kb, sequencing can be done after constructing a library.

9. Does BGI provide PCR amplification or product ligation?

   At the moment BGI does not provide PCR amplification. Product ligation must be performed by the client.

10. Can PCR product that was generated using universal primers be used in target region sequencing?

    We do not advise doing this. The presence of universal sequencing primers will cause light intensity fluctuation, resulting in an inaccurate adjustment of the machine. This will result in poor quality data.

11. When the fragment size of the PCR indexed product varies (e.g. 200-800bp), and only partial sequence information from both ends is needed, can high-throughput sequencing be applied?

    Two difficulties are inherent in this situation:

    Adapters should be ligated before sequencing, and the smaller fragments are prone to be ligated with the adapters, whereas the larger ones have a smaller ligation rate and thus a less effective concentration.

    A large sequence span (200-800bp) will lead to inhomogeneous signal quality in sequencing, resulting in poor quality data.

12. What are the services that BGI can provide for X chromosome sequencing?

    BGI offers the following X chromosome sequencing services:

Whole chromosome: To date, there has not been a very efficient method to capture the whole X chromosome. Sequencing can be done after library construction and amplification using chromosome segregation and micro-dissecting. However BGI does not offer this type of library construction and sequencing at the moment.

Targeted region on X chromosome: Information about the target region is required before sequencing and for designing of probes. Sequencing can be done following the usual target region sequencing.

Exon region on X chromosome: Agilent has an X chromosome exon reagent kit that can be used directly on the X chromosome for exon region target capture.