LGC Genomics GmbH
Ostendstrasse 25, 12459 Berlin, Germany
t +49 (0)30 5304 2200 **f** +49 (0)30 5304 2201
e info.de@lgcgroup.com **w** lgcgroup.com biosearchtech.com

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

Data required for designing oligo probes for a Flex-Seq assay

To start your project, the first information we require is the 'input data'. This refers to a file of target SNPs and a file of information on the corresponding reference genome.

1. A reference genome file (FASTA format) or details of the publicly available reference genome sequence to be used. Please provide us with a link for downloading, or provide the exact version of a genome reference sequence,
   e.g. Homo sapiens (assembly GRCh38.p11); hg38 will not suffice.
2. A file giving SNP/sequence variations locations (BED format); these locations refer to the given reference genome.

In more detail:

1. FASTA

A sequence file in FASTA format can contain several sequences. Each sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line must begin with a greater-than ('>') symbol in the first column. The word following the '>' symbol is the identifier of the sequence, and the rest of the line is the description (optional).

FASTA example:

```
>AB000263 |acc=AB000263|descr=Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.|len=368
ACAAGATGCCATTGTCCCCCGGCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCGCTGCCCTGCC
CCTGGxAGGGTGGCCCCACCGGCCGAGACAGCGAGCATATGCAGGAAGCGGCAGGAATAAGGAAAAGCAGC
CTCCTGACTTTCCTCGCTTGGTGGTTTGAGTGGACCTCCCAGGCCAGTGCCGGGCCCCTCATAGGAGAGG
AAGCTCGGGAGGTGGCCAGGCGGCAGGAAGGCGCACCCCCCCAGCAATCCGCGCGCCGGGACAGAATGCC
```

2. BED

1. **chrom** - The name of the chromosome (e.g. chr3, chrY, chr2_random) or scaffold (e.g. scaffold10671).
2. **chromStart** - The starting position of the **feature** in the chromosome or scaffold. The first base in a chromosome is **numbered 0**.
3. **chromEnd** - The ending position of the **feature** in the chromosome or scaffold. The *chromEnd* base is not included in the display of the feature. For example, the first 100 bases of a chromosome are defined as *chromStart=0*, *chromEnd=100*, and span the bases numbered 0-99.

BED example:

field 1	field 2	field 3	optional field 4
chr6	20921570	20921571	example_SNP_1
chr12	65783448	65783449	example_SNP_2
chr2	40692483	40692484	example_SNP_3

Depending on the species and the selection of SNPs, a certain number of SNPs will fail in the design phase. Therefore, we usually recommend including a minimum of 30% of additional SNPs for the first-round design. After the design is ready you can then select the number of SNPs you would like to keep according to your priorities and the specificity of the designed probes.

Please provide this information in the Sample Submission form by completing the tabs 'General' and 'Flex-Seq sample information' (mandatory fields). The remaining tabs will complete automatically. Once complete send it back by email to the Project Manager Next Generation Sequencing.

Project Manager Next Generation Sequencing
Anne Klingbeil/Sabine Nürnberg
Tel: +49 (0)30 5304 2268
Fax: +49 (0)30 5304 2201
Email: flexseqberlin@lgcgroup.com

LGC