~~HiSeq 3000, HiSeq 4000 (CSP ● Version 2 replaces HP10 in cBot Cluster Generation Reagent Plate)~~

~~Usage of a custom sequencing primer is currently not supported on HiSeq 3000 and 4000 machines. A work around as described for the HiSeq2500 (**CSP** ● Version 2 **REPLACES** HP10 in the cBot Cluster Generation Reagent Plate) is possible though. **ATTENTION:** Do not add the **CSP** ● Version 2 to the HP10 solution! A primer mixture would result in low clusters calls and the resulting reads would be contaminated by poly(T) stretches. Always use fresh HT1 and add the **CSP** ● Version 2 / HT1 dilution to the empty and rinsed well.~~

13. Appendix G: Data Analysis

This section describes a basic bioinformatics workflow for the analysis of QuantSeq data and is kept as general as possible for integration with your standard pipeline.
QuantSeq is available in two read orientations: QuantSeq FWD (Cat. No. 015) contains the Read 1 linker sequence in the 5′ part of the second strand synthesis primer, hence NGS reads are generated towards the poly(A) tail. To pinpoint the exact 3′ end, longer read lengths may be required. Read 1 directly reflects the mRNA sequence.
In QuantSeq REV (Cat. No. 016), the Read 1 linker sequence is located at the 5′ end of the oligodT primer. For Read 1, a Custom Sequencing Primer (included in the kit) has to be used. With QuantSeq REV it is possible to exactly pinpoint the 3′ end during Read 1. The reads generated here during Read 1 reflect the cDNA sequence, so they are in a strand orientation opposite to the genomic reference. For paired-end sequencing we strongly recommend using QuantSeq REV (Cat. No. 016).

For more detailed information please refer to https://www.lexogen.com/quantseq-data-analysis.

Processing Raw Reads

We recommend the use of a general fastq quality control tool such as FastQC or NGS QC Toolkit to examine the quality of the sequencing run. These tools can also identify over-represented sequences, which may optionally be removed from the dataset.

De-Multiplexing

External Barcodes: The barcode is contained in the Index Read, and demultiplexing can be carried out by the standard Illumina pipeline.

Trimming

As second strand synthesis is based on random priming, there may be a higher proportion of er-

rors at the first nucleotides of the insert due to non-specific hybridization of the random primer to the cDNA template. These mismatches can lead to a lower percentage of mappable reads when using a stringent aligner such as TopHat2, in which case it may be beneficial to trim these nucleotides. For QuantSeq FWD (Cat. No. 015) the first 12 nucleotides of Read 1 need to be removed. Alternatively, a less stringent aligner (e.g., STAR Aligner) could be used with the number of allowed mismatches being set to 14. While trimming the first nucleotides can decrease the number of reads of suitable length, the absolute number of mapping reads may increase due to the improved read quality. Reads, which are too short or have generally low quality scores should be removed from the set.

While single-read sequencing does not require any trimming using QuantSeq REV (Cat. No. 016), paired-end sequencing may require the first 12 nucleotides of Read 2 to be trimmed. Alternatively, also here the STAR Aligner could be used with the number of allowed mismatches being set to 16 for paired-end reads.

Alignment

At this point the filtered and trimmed reads can be aligned with a short read aligner to the reference genome.

STAR aligner or TopHat2 can be used for mapping **QuantSeq FWD** (Cat. No. 015) data. The reads may not land in the last exon and span a junction hence splice-aware aligners should be used. For **QuantSeq REV** (Cat. No. 016) we do not recommend using TopHat2, since there is hardly a need to search for junctions. Nearly all sequences will originate from the last exon and the 3' untranslated region (UTR). In case of no detected junction, TopHat2 may run into difficulties. Hence, Bowtie2 or BWA can be used for mapping in this case.

Annotation

Mapping only the 3' end of transcripts requires an annotation which covers the 3' untranslated region (UTR) accurately. Conservative annotations might decrease the power of correct gene quantification after mapping, especially in case of QuantSeq REV (Cat. No. 016). For some gene annotations it might be an advantage to extend the 3' UTR annotation further downstream in order to assign the mapped read correctly.

Please visit our website (www.lexogen.com) for an up-to-date table of suggested species-specific annotations and comments.

More information about the principal data analysis can be found under www.lexogen.com.