# Project report for BioClavis service project ({{ bcl\_project\_name }})

{{ customer\_name }} - {{ project }}

## Overview and purpose

This report describes the results (both QC and experimental samples) and links to results files generated from the TempO-Seq assays performed in this study. The work described in this report is a study using targeted sequencing-based RNA expression analysis of {{ sample\_type }}.

If you have any questions, contact BioClavis, either a member of the project team or support@bioclavis.co.uk.

## Sample information and plate layouts

Samples were supplied in {{ num\_plates\_slides }} \* {{ sample\_submission }} (received {{ samples\_received\_date }}), containing {{ num\_samples\_rec }} samples of which {{ num\_samples\_proc }} samples were to be processed.

The samples were processed, and were analysed using the {{ ref\_gen\_vers }} panel with standard attenuators.

## QA/QC test results

QA/QC of BioClavis’ internal process control data is comprised of the following steps, for which results are shown below. This is in addition to confirming success of the sequencing summary output file for each sequencing run to manufacturer’s spec containing the following metrics: Density (K/mm2), Clusters PF (%), Reads (M), Reads PF (M), and % ≥ Q30.

### Read count analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Value | Status | Significance |
| Number of mapped reads in positive RNA controls (average) | {{ avg\_mapped\_reads\_pos\_RNA }} | Pass (>{{ reads\_per\_sample }}million mapped reads) | Sufficient number of mapped reads are required |
| Signal:noise ratio  (total number of mapped reads in the positive controls / total number of mapped reads in negative controls) | {{ signal\_noise\_ratio }}:1 | Pass (>20:1) | Important quality parameter for a variety of statistical analyses. |
| The percentage of mapped reads in positive controls | {{ avg\_mapping\_perc\_pos}}% | Pass (>80%) | Ensures adequate yield, not a quality score as in RNA-Seq. |
| Average reads / probe (>20 reads) | {{ mean\_avg\_reads\_probes\_over\_20 }} | Pass (>250) | A measure of ‘read depth’ that demonstrates the number of times expressed genes were counted. |

### Replicate analysis of positive RNA controls and no-sample negative controls

Process controls are run in replicate on each assay plate of samples to ensure quality metrics pass on a plate-wise level. Representative scatter plots of log2 transformed read counts from process control replicates are shown below. The assay in this project generated high quality data, as demonstrated by the high reproducibility of the positive controls, and the low signal in the negative control.

{{ pos\_con\_fig }} {{ neg\_con\_fig }}

### Samples PCA Plot

Although the standard QC is primarily based on the process controls added while running the assay, overall performance of the experimental samples was briefly examined. This was a quality check and was not intended to be definitive or exhaustive. In the PCA plot below, samples and controls cluster in a generally expected way, so do not indicate any quality concerns.

{{ pca\_fig }}

## Data files

### Counts per gene per sample

A data matrix of gene expression level (raw counts) with sample names as column headers and gene names as rows is the primary output of this study. The data file ‘{{ data\_file }}’ will be transferred electronically.

### Attenuation

Note, some commonly high expressed genes have been attenuated in this assay resulting in their counts being reduced in a linear and predictable manner. The list of these attenuated probes is included in the manifest. The results with attenuated counts are perfectly valid for fold difference calculations, since attenuation impacts all samples the same. However, if you would like a data file with these genes counts ‘back calculated’ to what they would have been, for example in comparing absolute counts across different platforms, we can provide these. Note, this calculation requires normalisation of the dataset.

### Normalised counts per gene per sample

The data provided have not been normalised, as selecting the appropriate method of normalisation is dependent on downstream analyses.

### Sequencing output

Raw sequencing data files in FASTQ format will be delivered electronically, if requested.

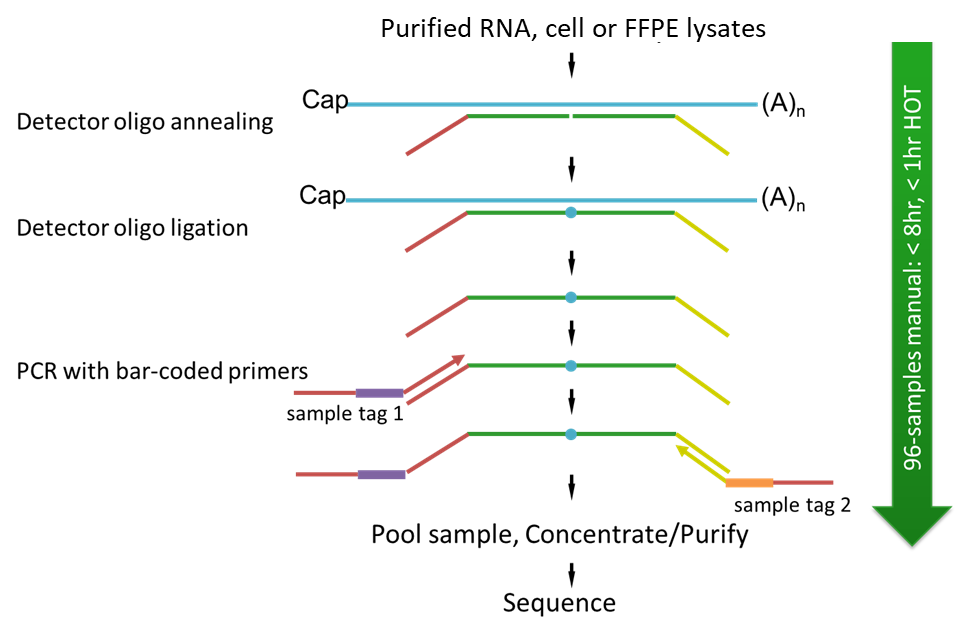
### Protocols

#### TempO-Seq assay protocol summary

Sequencing libraries for targeted panels were generated as described briefly below and depicted in the Figure.

In TempO-Seq, each Detector Oligo consists of a sequence complementary to an mRNA target plus a universal (i.e. same for every targeted gene) primer binding site. They anneal in immediate juxtaposition to each other on the targeted RNA template such that they can be ligated together. Ligated detector oligos are PCR-amplified using a primer set (single-plex PCR reaction, with a single primer pair for each sample) that introduces both the adaptors required for sequencing and a sample-specific barcode. The barcode sequences flank the target sequence and are inserted appropriately into the standard Illumina adaptors to permit standard dual-index sequencing of the barcodes and deconvolution of sample-specific reads from the sequencing data using the standard Illumina software. All the PCR-amplified and barcoded samples are pooled into a single library for sequencing. Sequencing reads are demultiplexed using the standard sequencing instrument software for each sample using the barcodes to give a FASTQ file for each.

Figure: Schematic of TempO-Seq assay protocol



#### Data analysis protocol

TempO-Seq sequence files were analysed using the Tempo-SeqR software package. The input for TempO-Seq data analysis is a folder of zipped FASTQ files. Each FASTQ file contains the reads and quality scores for one sample. Each FASTQ file is aligned using the STAR algorithm to a pseudo-transcriptome corresponding to the gene panel used in the assay. It takes ~3 minutes to align 1 million reads using the Tempo-SeqR software. The primary output of the Tempo-SeqR software is a table of counts with each column representing a sample and each row representing a gene.

#### Data normalisation

Data normalisation is commonly not necessary. A variety of normalisation methods may be appropriate for various downstream analyses at your discretion, including normalisation to total reads per sample.

#### Probe manifest

ProbeID is composed of gene name along with a numeric identifier, separated by an underscore. Some genes may have multiple ProbeIDs representing assays against isoforms that are not captured by a single probe against the region of maximal overlap of all gene isoforms. For additional information about probe location, including target sequence and RefSeq ID, refer to the Probe Manifest file that will be delivered electronically.

## Return of remaining materials

If you would like the remaining RNA and libraries returned at your cost, please request shipment within 3 months unless arrangements were already made. You will need to supply a QuickSTAT (or other courier) account number to charge against. We cannot guarantee that any materials are left or warrant their quality but will attempt to satisfy requests as much as possible.