Description of Xpress-seq bulk RNA-seq data deliveries

This document describes the content and format of Xpress-seq bulk RNA-seq preprocessing data output and delivery.

Overview

In short, the data the customer will receive in return after completed library service consist of the following:

- Raw Data (demultiplexed)
- Aligned Reads
- Count Tables
- FastQC output
- Metadata and other files

Raw Data

Raw sequencing data is provided in gzipped fastq format, demultiplexed per sample. Depending on the ordered read depth and number of samples, the raw data may consist of partial, full or multiple flow cells.

Furthermore, fastq files are provided after trimming low quality bases and adapters.

Aligned Reads

We utilize STAR (Dobin et al., 2013) to align trimmed reads to the reference genome (hg38 for Human / mm38 for Mouse).
The resulting output is coordinate sorted, indexed read files in bam format. Unmapped reads are retained in the bam file.

Count Tables

Gene expression values are determined from aligned reads using featureCounts (Rsubread package) using gencode gene annotations. Gene expression count tables are provided in plain text format for convenience. We provide two quantifications:

a. Exon-mapped reads
b. Intron+Exon-mapped reads

All count matrices have cells in columns and genes in rows of the table.

Postal address
Xpress Genomics AB
Visionsgatan 56
171 64 SOLNA
Sweden
Org. number 559400-8202

Visiting address
Visionsgatan 56
171 64 Solna, Sweden

Telephone
+46 760079794

VAT Number
SE559400820201

E-Mail
hi@xpress-genomics.com
Web
www.xpress-genomics.com

FastQC output

We provide a per-sample output of FastQC analysis containing read numbers and sequencing quality metrics. Furthermore, these are also aggregated into a single MultiQC report (html file to be viewed in browser).

Metadata and other files

In addition to the previous files and outputs, data submissions contain the following:

- Statistics file from gene assignment using featureCounts, indicating the number of reads falling into different assignment and mapping categories.
- Optional: If sample names are not used as file names, we provide a mapping table of sample barcodes to sample names.
- Text file providing md5 checksums to confirm integrity of the data transfer.

For questions, please contact us at service@xpress-genomics.com.