umcg
Genetica
Genome Analysis Facility

Title:	Guidelines Analysis of RNA Quantity and Quality for Next-Generation Sequencing Projects
Protocol number:	GAF S002
Version:	Version 4
Date:	December 17th 2015
Author:	P. van der Vlies, C.C. van Diemen

Guidelines Analysis of RNA Quantity and Quality for Next Generation Sequencing Projects	GAF S002

Introduction

To ensure good quality data and a successful and rapid processing of total RNA samples for transcriptome sequencing, there are a number of points to consider. Also the RNA quality has to fulfil a number of requirements to be suitable for use. Properties such as purity, integrity and concentration are of great importance, because they can have a significant impact on the experiment's results.
All the required information before start of an RNAseq project performed by the Genome Analysis Facility of the UMCG (GAF) is described in this document.

Applications and limitations of the technique

With RNAseq, cDNA is being sequenced and provides information on differential expression of genes, including gene alleles and differently spliced transcripts; non-coding RNAs; post-transcriptional mutations or editing; and gene fusions. These data can be used to answers multiple research questions such as:

- Identification of **genome-wide differential expression of genes** across two or more conditions with statistical significance and/or biological significance (e.g. fold change)
- **Pathway analysis** (a.k.a. gene set analysis) to infer correlation of differential expression evidence in the data with pathway knowledge from established databases
- **Co-expression analysis** and network analysis of RNAseq data are used to investigate potential transcriptional co-regulation and gene interactions
- **Prediction analysis** for clinical practice

Points to consider prior to the start of the experiment

In design of the study and data analyses:

- The design of the RNAseq study and the analysis of RNAseq data is rather complex. The primary data analysis is supported by the GAF and Genomics Coordination Center (GCC). However the applicant remains responsible for the final analysis. It is a complicated technology, so ensure sufficient knowledge, computational skills and computer power.
- Resource limitation vs. statistical power: ensure sufficient samples to allow robust results
- The technique is meant for genome-wide expression analysis, not to analyse only a few genes. qPCR is then better suited.
- Replication problems between studies may occur due to diversity of study designs
- Cellular diversity within samples: expression variation is largely within or among cells
- Ensure sufficient sequence depth for analysis of differential expression of low expressed genes and novel splice variants detection, when desired. The GAF will generate standard 15 million reads or read pairs on avarage for each sample.

- In case of the use of a standard (poly A) mRNA sample prep kit, only poly-A tailed RNA transcripts are sequenced. In this case contamination with e.g. abundance of globin mRNAs in whole blood, significantly influence the whole gene expression profile
- Normalization of the data is necessary
- No consensus on statistical thresholds for differential expression
  - fold-change
  - Bonferroni correction
  - false discovery rate

Practical aspects:

- The technology is sensitive for batch effects: avoid clustering of samples due to batch differences or technical effects: samples should be randomized and treated in the same way as much as possible.
- The GAF is able to perform several library preparation methods en sequencing methods. See table for validated methods.

Library Preparation method	**Sequencing method***	**Target amount of reads/ readpairs***	**Description**
Lexogen QuantSeq 3' mRNA	• HiSeq 2500: 50 bp Single Read • NextSeq500: 75 bp Single Read	4 million	• Reads to 3' end of the transcript • for Gene Counting • Stranded • cost effective
Illumina TruSeq mRNA (polyA)	• HiSeq 2500: 50 bp Single Read or 2 x 100 bp Paired End • NextSeq500: 75 bp Single Read or 2 x 150 bp Paired End	15 million	• Entire poly A transcript • For Gene Counting and SNP detection • Non-Stranded
Illumina TruSeq RiboZero	• HiSeq 2500: 2 x 100 bp Paired End • NextSeq500: or 2 x 150 bp Paired End	15 million	• Entire transcript • rRNA depletion: yield polyA transcripts and non-conding RNA's • For Gene Counting and SNP detection • Stranded

*Adaptation of sequencing method and target amount of reads is possible after discussion with the GAF

- For RNAseq experiments the GAF guarantees a successful experiment for samples containing enough sufficient pure and non-degraded RNA. The GAF has a high success rate in case for samples with fewer starting material, however a successful experiment for these samples is not guaranteed. The complexity of the generated reads could be reduced in samples with limited starting material. See also the section “RNA quality and integrity”in this document
- If possible, the samples are processed in an automated fashion using the Perkin Elmer SciClone liquid handler in series of 8-96 samples. Series of 96 samples are most cost-effective.
- After finishing the sequence run, demultiplexing of the obtained reads, alignment and SNP detection is performed by Bio-informaticians of the Genome Coordination Centre GCC. Final data are delivered by the GAF.

RNA isolation

RNA isolated by most isolation methods usually meets the criteria for a successful next-generation sequencing experiments. The RNA should be dissolved in 50 µl Elution buffer or 10 mM Tris-HCL pH 8.0, 1 mM EDTA. We have successfully processed RNA obtained by the Qiagen RNeasy-kit, Norgen kits or Trizol reagent.
Treat all samples in the same way as much as possible to avoid batch-effects.

RNA quantity and quality

1. The quantity and purity of the isolated RNA should be initially verified with a spectrophotometer, ideally by the Nanodrop spectrophotometer.
   The measurement of the integrity of the RNA and a more precise measurement of the concentration should be performed using capillary electrophoresis device. (Bio-Rad Experion Bioanalyzer, Agilent Bio-Analyzer or PerkinElmer LabChip GX)
2. As extra service, the GAF can perform the Nanodrop measurement and Capillary electrophoresis for RNA samples. This is done in a weekly routine according to document GAF-ALG005.
3. The GAF quantity and purity criteria for RNAseq are:
   a. The OD 260/280 ratio should be above 2.0, the OD 260/230 ratio should be above 2.0 (Nanodrop measurement)
   b. RQI/RIN-score is > 7.0. (measurement with capillary electrophoresis)
   c. Concentrations and volumes according table

Library Preparation method	**Concentration (based on capillary electrophoresis)**	**Volume**	**Remarks**
Lexogen QuantSeq 3' mRNA	2-100 ng/µl	> 5 µl	Concentrations should be normalized as far as possible
Illumina TruSeq mRNA (polyA)	10-20 ng/µl	> 50 µl	
Illumina TruSeq RiboZero	20-100 ng/µl	> 10 µl	

Preparing your samples and shipping them

1. The RNA solution should be transferred to an *RNAse free* 96-well plates (for instance: Greiner Bio-One, PCR-plate full skirted, order nr. 652270).
2. To avoid seeing batch differences in the final data, it is important to put the samples in the plate in a random order, starting with position A1, then B1, etc. Do not leave any blank wells between samples.
3. The 96-well plates should be firmly closed, preferably with a Thermo-seal (*Thermo scientific order no. AB-0559*) using a heat-plate sealer. Alternatively, if a heat-plate sealer is not available, adhesive PCR foil seals (*Thermo scientific order no. AB-0626*) can be used.
4. Record the sample IDs in a special XLS file according to the instructions below. The GAF doesn't accept samples with names of individuals. A unique plate name should be given to each 96-well plate, using a permanent marker to annotate each plate on both sides.
5. The plates should be stored and transported at –80°C. Deliver the samples on dry ice to the Genome Analysis Facility lab or send your samples by courier to the Genome Analysis Facility

laboratory at the Department Genetics, UMCG. The shipment should be placed in boxes containing a sufficient amount of dry ice.

6. To avoid any delay during the weekend, please ship the material at the beginning of the week.
7. The Genome Analysis Facility is not responsible for long time storage of samples or intermediate products. Left overs are thrown away after finishing the project.

Documenting the sample and plate layout

The administration of all sample IDs and their corresponding wells in the 96-well plate should be recorded in an XSL file (Excel) (see figure 1). This XSL file should summarize all the required information:

- Number
- Plate name
- Well with corresponding unique sample ID
- Information on the RNA quality: RNA concentration, OD 260/280 ratio, OD 260/230 ratio.

nr	Plate	Well	SampleID	RNA-concentration (ng/ul)	OD260/280	OD 260/230	Volume (ul)	RIN/RQI value
1	UMCG-1	A1	345	71	2.12	2.39	20	8
2	UMCG-1	B1	456	82	2.07	2.4	20	8
3	UMCG-1	C1	564	56	2.12	2.38	20	9
4	UMCG-1	D1	873	65	2.09	2.28	30	10
5	UMCG-1	E1	937	87	2.1	2.34	30	8
6	UMCG-1	F1	345	93	2.06	2.23	20	9
7	UMCG-1	G1	244	52	2.11	2.43	20	10
8	UMCG-1	H1	879	79	2.12	2.19	25	8
9	UMCG-1	A1	435	80				
	UMCG-1	B2						

Figure 1. Example of the attached XSL file

Please email the XLS file with your sample information to: p.van.der.vlies@umcg.nl

Delivery address

Genome Analysis Facility
ERIBA, 5th floor
Dept. Genetics, UMCG CB53
Hanzeplein 1
9713 GZ Groningen
The Netherlands

Phone number: +31 (0)50-3617100
Email: p.van.der.vlies@umcg.nl

Before submitting your samples, please contact the Genome Analysis Facility