RNA submission guidelines for RNA-seq

How to submit your RNA samples?

- Fill in the **RNA submission excel file** found on our website and email it to us. This must contain all relevant information, no information can be transmitted only within the email text or orally.

- Provide the **RNA profile analysis** if available. If you haven't done one, we will do it before starting the prep. Since this can induce some extra delays in case of inconclusive results, we strongly suggest you perform it before submitting your RNAs in order to streamline the whole workflow. We can train you on our TapeStation instrument, which is very user friendly and easy to operate. Request a training 1-2 weeks before your RNA extraction.

- If you submit several **independent projects**, use different submission forms. In case these projects should still be sequenced together, indicate it in the comments.

- Bring RNAs on dry ice, in RNase-free **PCR-grade 0.5ml-2ml microtubes**, but not in PCR tubes/strips/PCR plates. Leave space on the side of tubes for our internal codes. Do not use tape as it detaches at -80°.

- Indicated tube **labels** must match exactly the ones on your physical tubes. If available, names of profiles must also match perfectly either tube labels or sample names.

- Avoid using "S1", "S2"… in your samples name. These codes are already appended by the sequencer to fastq files, which could cause confusion.

- If you need to **modify** or comment your submission in any way, do it by email by resubmitting an updated file, not orally.

Recommendations for RNA preparation

- **RNA quality and quantity** are crucial factors for the outcome of the RNA-seq. There are methods starting from tiny amount of poor-quality RNA, but these will give lower quality data. Therefore, we strongly recommend that you optimize RNA extraction before the real experiment.

- For a differential expression analysis, we recommend at least **triplicate biological replicates** in order to generate sufficient statistical power.

- Avoid **batch effects** as much as possible when designing your experiment: distinct biological groups should not be processed in distinct batches during samples preparation/RNA extraction.

- If you are interested in **small RNAs**, use a suitable RNA extraction method (not all are).

- Avoid **bacterial contamination**, for instance during dissection or FACS sorting (e.g. using old buffers).

- **Degraded RNAs** are not a suitable starting material for most of the key library prep methods, therefore follow good RNA-specific lab procedures when preparing/handling RNAs.

- **Extraction methods** finishing with an affinity column and washes (such as RNeasy…) yield RNA of better quality than methods finishing with precipitation/resuspension (such as Trizol…). If you still prefer using a

Trizol/Trizol-like method, consider coupling it at its final steps with a column-based method, for instance the Zymo "RNA Clean and Concentrator-5 kit with DNase" kit, following manufacturer's instructions.

- Include a **genomic DNA elimination step**, such as:

- on-column DNase treatment during RNA extraction (e.g. Qiagen DNeasy with the optional DNase)
- genomic DNA exclusion column during RNA extraction (e.g. Qiagen DNeasy PLUS kit)
- on-column DNase treatment performed after RNA extraction (e.g. with Zymo RNA Clean and Concentrator-5 kit with DNase, #R1013 or R1014). *Useful for Trizol or Trizol-like extraction methods.*

Note: A simple post-extraction DNase treatment without subsequent column purification can cause major downstream problems so absolutely avoid doing that.

Note: Standard TapeStation or Fragment Analyzer runs do not detect contamination with intact gDNA.

- **RNA elution** should be done in RNase-free H2O or Tris, but not in a buffer containing EDTA such as TE, which would inhibit some downstream applications.

- Nanodrop measurements of **RNA concentration** may be very inaccurate, in particular below 50 ng/ul where extraction buffer contaminations can be responsible for most of the absorbance. Therefore, report 260/280 and 260/230 ratios when nanodrop is used, and if possible use the "nanodrop One" instrument which corrects for most buffers contamination (only above 20 ng/ul though). Alternatively, a more accurate buffer-insensitive quantification method is qubit RNA HS.

- With good quality RNA, there is no lower limit to the **RNA quantity** required for RNA-seq, but more RNA will give more sensitivity and accuracy. In more details:

- The dream world quantity to submit is 2 ug RNA in max 50 ul.
- Above 100ng (in max 50ul) we can still use our gold standard stranded method. This is clearly what you should aim for as a minimum, samples allowing.
- Above 2ng total (in 20ul max) we can use with good results the smart-seq v4 method, even though it is non-stranded and more expensive.
- Below that there is no threshold anymore, we would use the smart-seq v4 as well but with decreasing results quality.

- **RNA storage** must be done at -80°C.

Determining RNA-seq library prep method

The choice of a library prep method depends directly on results from RNA quantification and QC, and is a shared decision between the user and the GECF. This discussion will address the following topics:

- Are you interested only in mRNAs or also in **lncRNAs**?
- Is detection of **small RNAs** (miRNAs…) required?
- If many samples are to be submitted (such as >40), would a "**3' end method**" be suitable? Such methods, which conserve only the 3'end of mRNAs, allow for cheaper library prep and sequencing.

- Will you need **long reads**, such as paired-end 150nt reads (2x 150nt)? If yes, libraries protocols will be tweaked accordingly. Only certain specific RNA-seq projects benefit from such read length, notably for repetitive transcripts such as retrotransposons, or for quantifying splicing isoforms.
- Are your RNAs likely to contain **bacterial RNAs** (gut samples, neutrophils fighting an infection…)? If yes, do they need to be sequenced?
- **Strand information**, which helps reads mapping (on overlapping genes of opposite orientation, or in presence of genomic DNA contamination) is provided by most methods. Yet, for low quantity/quality RNA, we use a "non-stranded" method, please tell us in advance if this is an issue.
- Is the total RNA content as observed by nanodrop/qubit/TapeStation/BioaAnalyzer misleading in terms of the **actual amount of RNA molecules of interest**? Typical cases are: 1) nuclear fraction was isolated (rRNA is lower, thus proportionally mRNA is higher); 2) fractionation of lysates into polysomes and other ribosomal-related fractions (some fractions may not contain mRNA).

Sequencing settings

Defining sequencing settings is not strictly required at RNA submission stage, but can already be discussed with us or with the bioinformatician that will perform the analysis.

Sequencing settings depend on the aim of your project, such as global gene expression profiling, detailed quantification of rare transcripts, quantification of distinct splicing isoforms, identification of the sequence of new transcripts… A few questions you can already think about:

- Do you need single-end or paired-end reads?
- Which reads length do you need? (see above)
- How many reads per sample do you need? Recommendations range from 15mio (low resolution broad transcriptional profiling), to 25-40mio (standard transcriptional profiling), and more if SNPs are investigated.
- In case the number of samples dictates to perform several sequencing runs, how would you like your samples to be split across runs to minimize batch effect?

Disclaimers

- For standard libraries prepared by the GECF, we guarantee a reads quantity with a 20% margin. For instance, for a NextSeq run of 400 mio reads containing 10 samples, we guarantee 32 mio reads/sample. When libraries prepared with different methods are sequenced together, no firm guarantee can be given.

- The GECF guarantees successful library preparation when using a commercial method starting with RNAs with QC matching that method's specifications. In case a library prep does not lead to satisfactory results due to technical problems, the GECF will prepare it again without charges. Yet, when it turns out that the issue does not come from the GECF (e.g. problem is reproducible, or a control library is successfully prepared in parallel), the libraries prep will be billed to user. Typical, but very rare, examples of GECF-independent issues are:

- Library prep is inhibited by a contaminant present in the RNA samples, typically a buffer (Trizol, guanidine salts) or genomic DNA, enzymes… These can be difficult to detect.
- A wrong RNA concentration was indicated by user. In case of suspicion of a quantification issue and when sample quantities allow, we independently check user-provided concentrations.

- When the GECF and users agree to perform an exploratory method, or to process samples even though they do not comply with recommended specs for the chosen method, the GECF cannot be held responsible for failures or unsatisfactory results (outside errors from the GECF side).

- The GECF is an open access facility, therefore we cannot give guarantee regarding the confidentiality of data. If this is an issue, the easiest solution is to give codes to your sample.

Versions log

- v1.01, 1.02: initial/old releases.

- v1.03: Indicated that RNA quality and quantity is a crucially important factor for the outcome of the RNA-seq experiment. Clarified to avoid using "S1", "S2"… in your samples name. Clarified that the *Nanodrop One* correction of buffer contamination only works above 20 ng/ul. Mentioned cases where total RNA content may be misleading for downstream protocols (nuclear fraction, polysomes…).

- v1.04: Indicated not to use tape to label RNA. Modified RNA quantities to submit to comply with new methods used at GECF.

- v1.05: Clarified that the submission form itself must contain all information. Clarified that providing the profile analysis will streamline the workflow and likely shorten turnaround time.