guidelines v1.03

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.

- 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 some space on the side of tubes for our internal codes.

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

- Avoid using "S1", "S2"… in your samples name. These codes are indeed 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, even when you have already told us orally.

Recommendations for RNA preparation

- CAUTION: RNA quality and quantity is a crucially important factor for the outcome of the RNA-seq experiment. There are methods allowing starting from tiny amount of poor quality RNA, but these will systematically give lower quality results than methods starting from higher quantity of good quality RNA. Therefore, we strongly recommend that you optimize RNA extraction before the real experiment, by testing different kits/reagents and tissues disruption methods.

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

- Avoid **batch effects** 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 (if using old contaminated 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. with Qiagen DNeasy kit with the optional DNase)
  - genomic DNA exclusion column during RNA extraction (e.g. with 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).

  Note: Trizol or Trizol-like extraction methods (typically non-column based) can still incorporate an on-column DNase treatment during extraction, by coupling them with a column method (see comment above).

  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 <u>do not</u> 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 at low concentrations (<30 ng/ul) where putative extraction buffer contaminations can be responsible for most of the absorbance. Therefore, always report 260/280 and 260/230 ratios when nanodrop is used, and if possible use the "nanodrop One" instrument which detects most usual buffers contamination and corrects values accordingly (only when above 20 ng/ul though). Alternatively, a more accurate buffer-insensitive quantification method is qubit RNA HS.

- There is no lower limit to the **RNA quantity** required for RNA-seq, but more RNA will clearly give better results (more sensitivity and accuracy). In more details:

  - The dream world quantity to submit is 8 ug RNA in not more than 100 ul. This allows to perform the RNA-seq in best possible conditions.
  - Above 500ng (in max 100ul) we can still use with nice results the gold standard TruSeq stranded mRNA 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 than truseq stranded mRNA.
  - Below that there is no threshold anymore, we would use the smart-seq v4 as well but with decrease results quality.

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

- RNA quality and quantity are important determinants of the quality of the subsequent RNA-seq. We thus recommend that, when working with new and possibly tricky tissues/settings, you perform a **trial extraction and QC** before starting the real RNA experiment.

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, which you can already describe when inquiring by email:

- 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 with longer fragments will have to be generated, and protocols will be tweaked accordingly. Only certain specific RNA-seq projects really benefit from such read length, notably studies of repetitive transcripts such as retrotransposons, or studies quantifying specific splicing isoforms.
- Are your RNAs likely to contain **bacterial RNAs** (gut samples, neutrophils fighting an infection…)? If yes, and if a non-polyA-based method has to be used (typically the case if RNAs are degraded), the library prep should include a bacterial rRNA depletion step. New methods less sensitive to RNA degradation (since keeping only the 3' end of mRNAs) and to bacterial contamination (since poly-A based) could also be envisioned.
- Of note, **strand information**, which helps reads mapping (on overlapping genes of opposite orientation, or in presence of genomic DNA contamination) is a nice asset, and is provided by most methods. Yet, for low quantity/quality RNA, we may have to choose a "non-stranded" method, please tell us in advance if this is a major 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 (typically some may contain more or less mRNA).

Sequencing settings

Defining exact 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 RNA-seq 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?
- How many reads per sample do you need? 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. Of note, when libraries prepared with different methods are sequenced together, no guarantee can be given, but we will do our best to equilibrate the numbers of reads obtained by each library type.

- 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

- The GECF guarantees successful library preparation when using a commercial method starting with RNAs of concentrations and quality matching that method's specifications. In case a library prep fails or does not lead to satisfactory results due to technical problems, the GECF will prepare it again without additional charges. Yet, when it turns out that the issue does not come from the GECF (e.g. problem is reproducible, and/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 and can potentially harm library preparation.
- A wrong RNA concentration was indicated by user. Of note, in case of suspicion of a quantification issue and when sample quantities allow, we independently check user-provided concentrations. Yet sometimes RNA amount does not allow for this double-checking.

- 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 obvious 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: initial release.

- v1.02: Clarified that when submitting several independent projects, different submission forms should be used. In case these projects should still be sequenced together, it should be indicated as a comment. Also indicated that tube labels must match exactly the ones on your physical tubes. If available, labels of profiles must also match perfectly either tube labels or sample names. Clarified that when one needs to modify or comment a submission in any way, it must be done by email, even when the information has already been given orally.

- v1.03: Indicated not to use tape to label RNA samples since it can detach at -80°. Indicated that RNA quality and quantity is a crucially important factor for the outcome of the RNA-seq experiment. There are methods allowing starting from tiny amount of poor quality RNA, but these will systematically give lower quality results than methods starting from higher quantity of good quality RNA. Therefore, we strongly recommend that you optimize RNA extraction before the real experiment, by testing different kits/reagents and tissues disruption methods.". Described in more details the RNA starting amounts. Clarified to avoid using "S1", "S2"… in your samples name, since these codes are already appended by the sequencer to fastq files, which could cause confusion. 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…).