Libraries submission guidelines for NGS

Submission process

Users should fill in the excel submission form found on GECF website, and submit it by email to gecf@epfl.ch together with the profiles. Fields highlighted in red in the form are mandatory.

Libraries must have been cleaned up before submission with beads or columns. For elution, H2O, Tris or TE are acceptable.

Submit your tubes in a clear bag (provided on top of our fridge), labeled with your name, lab PI, short description and date. Place the bag in submission drawer of our -20°C. Once we have received both the emailed submission form and the samples, the project enters the queue. In case of significant issues in the form, profiles or tubes labels, the submission will go down in the queue until the issues are fixed.

Preferred tubes are 0.5ml microtubes or larger. Strips are accepted only for large number of samples.

If you need to modify or comment your submission, do it by email even when you have already told us in person.

Never take back a tube from that drawer, or swap a tube, without notifying us in advance.

The GECF is an open access facility, therefore we cannot guarantee the confidentiality of data. If this is an issue, the easiest solution is to code your samples names.

Libraries amounts to submit

The minimum amount for each library (of standard size) is >7ul at >1ng/ul, or >4nM by qPCR, but larger volumes and higher concentrations are preferred, and sometimes required (see below). *If you cannot reach these values, we will apply when possible a special protocol. Yet, this may reflect an underlying issue, often leading to low yield/high duplication rates and thus no guarantee of reads output.*

In addition, the minimal total volume to submit for the full set of libraries depends on the instrument (again more volume is better):

	Minimal total volume for the whole set (more better)
MiSeq	7ul
NextSeq	7ul
NovaSeq 400mio	16ul
NovaSeq 800mio	55ul*
NovaSeq 1'600mio	55ul*
NovaSeq 4'100mio	90ul*
NovaSeq 10'000mio	200ul*

* If these values cannot be reached, we can discuss an alternative workflow for an additional fee.

A few practical examples, considering libraries at >1ng/ul:

NextSeq with 1 lib: >7ul of this library
    "   "   10 lib: >7ul of each library

NovaSeq 400mio with 1 lib: >16ul of this library
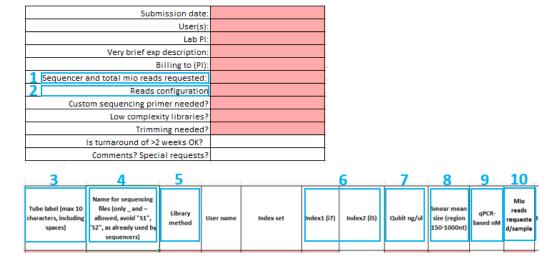   "   "   "   2 lib: >8ul of each library
    "   "   "   10 lib: >7ul of each library

NovaSeq 1'600mio with 1 lib: >55ul of this library
    "   "   "   5 lib: >11ul of each library
    "   "   "   10 lib: >7ul of each library

These examples assume equal distribution of reads across libraries. The volumes will need to be adjusted in case of very uneven distribution.

Additional precisions regarding fields in the form:

Submission date:	
User(s):	
Lab PI:	
Very brief exp description:	
Billing to (PI):	
1 Sequencer and total mio reads requested:	
2 Reads configuration	
Custom sequencing primer needed?	
Low complexity libraries?	
Trimming needed?	
Is turnaround of >2 weeks OK?	
Comments? Special requests?	

3	4	5			6		7	8	9	10
Tube label (max 10 characters, including spaces)	Name for sequencing files (only _ and – allowed, avoid "S1", "S2", as already used by sequencers)	Library method	User name	Index set	Index1 (i7)	Index2 (i5)	Qubit ng/ul	Smear mean size (region 150-1000nt)	qPCR-based nM	Mio reads requested/sample

1. Sequencer and total mio reads requested

Read configurations options change depending on the sequencer and mio reads selected. Sometimes the choice between nextseq and novaseq will require discussions, in such case choose "TBD".

Users can ask that we perform a **spiking** of a library into another run, when it is a non-standard library type and a significant chance exists that sequencing fails. When requesting a spike:

- Avoid absolutely mistakes of indexes (see comments in index section).
- Let us know if you plan to run it later in a full run, to make sure we will still have enough.
- If you have several samples to spike, let us know if they can be sequenced in distinct runs.
- Spikes cannot be allocated more than 8 mio reads in total. If you need significantly more than that, please rather consider a NextSeq mid-output run or a MiSeq run.

2. Reads configuration

For paired-end reads, the selected reads length should be consistent with the mean size of your libraries to avoid overlapping reads. Remember that your mean library size comprises ca 140bp of adapters.

The quality of the ends of 150nt reads decreases significantly after 130nt, so depending on your aims you may want to use only the first 130nt of the reads in downstream analyses.

Whenever possible (depending on parameters such library method, indexes…), we generate reads that are a little longer than specified by Illumina, for instance 80-85nt instead of the 75nt mentioned in our list price. If you plan to compare fastq files over large series of runs, you may want us to stick to the nominal read length for consistency (for instance 75nt). Ask us to do so in the comments section, if needed.

When requesting a spike, indicate minimal reads configuration requirements.

3. Tube label

Tube labels must be unambiguous, (ideally) max 10 characters, and perfectly match the ones indicated in the excel sheet. Leave some space on the side of tubes for our internal codes.

4. Sample name

Do not use "S1", "S2"…, nor "R1" or "R2"… in your samples name, as these codes are already appended to fastq files by the sequencers.

Only special symbols tolerated by the sequencer are – and _ .

5. Library method

Indicate the library preparation method/kit. If using a new custom technique for library prep, tell us and we may first do a spike. Without that, we cannot give guarantees in case of run failure, as the profile and quantity may look fine but the library structure may be faulty. A qPCR is also useful (see below).

6. Indexes

Indicate unambiguous index names (original names when commercial).

If <=2 samples are sequenced, base diversity of index reads may be too low. Contact us to discuss which indexes should be picked up. In particular, in such cases, avoid using Nextera indexes N705/N718, truseq LT index 11, NEBNext index 11, custom index Ad2.5, SureSelect XT indexes B05/F05.

When mixing libraries indexed with different sets of indexes, you can use our homemade Tindex website tool to check compatibility: https://gecftools.epfl.ch/tindex . Ask us in case of doubt.

NovaSeq (as HiSeq4000 previously) is subject to index hopping issues (indexes "jumping" from a fragment to another in the libraries pool). To alleviate this issue, which can be seen in a few percent of reads in bad cases, use only "UDI" indexes. In addition, remove free adapters (sharp peaks below 100nt).

IMPORTANT: it is crucial that users requesting spikes are 100% sure of the indexes they used. In case of error and thus indexes overlap, the main run may be destroyed. In such case, the spiking user will be

charged for the whole run. In the same manner, in case the main user announces a wrong index, its library may be affected by a spike. In that case, the main user is responsible and no discount will be granted.

7. Qubit values

To get accurate and consistent qubit values, we recommend: using the "1x dsDNA HS" kit; measuring all samples in parallel; running regularly new standard curves; and measuring regularly a control library.

In case you use other methods (dsDNA BR, picogreen, etc.), please specify it in the comments.

Provided enough volume is available, we'll check qubit values, and make a mean of both measurements.

We cannot be held responsible if a run fails due to inaccurate quantification values provided by a user.

8. Profiles analysis

Submit the pdf with mean smear size in a region of 150-1'000nt (we can help).

Profiles labels in the pdf must match the library names or tubes labels in excel file.

We have no way to double-check raw profiles when they were not performed at GECF, so we'll have to blindly trust the values given to us in that case.

If only an agarose gel is available, indicate the average size of the library and submit a properly labeled image (including ladder sizes). Gels are not recommended as they often fail to detect bands below 200nt.

Profiles should meet following criteria:

- no adapter dimers (120-150nt peak). Small amounts and in no more than 20% of the samples can be acceptable, but will still lead to loss of informative reads. An additional bead selection will help.
- main part of the smear should be below 800bp (if not, discussion is needed, as qPCR quantification and spiking may be necessary). Fragments bigger than that will be sequenced only inefficiently and may lead to overall decrease in run quality.
- smear should be bell-shaped unless specific case
- on the nextseq and miseq only, reasonable amounts of free adapters (<80nt) are not an issue.

9. qPCR-based nM

Certain complex custom libraries (HiC, 4C-UMI…) should be quantified by qPCR to get consistent results. Results are often good-enough without qPCR quantification, but we cannot give guarantees. In some extreme cases, the qPCR does not perfectly reflect the behavior in the sequencer, in these cases we cannot be held responsible for issues of reads output.

Ask us for a protocol if needed. We recommend the JetSeq Library Quantification Lo-ROX Kit.

10. Mio reads requested

When libraries were prepared by GECF, we guarantee a read output per sample of at least 80% of the anticipated number of reads.

When running a custom library type, no number of reads can be guaranteed. We do our best to give you satisfactory results though.

When mixing different libraries types, it is difficult to predict reads output accurately due to their different sequencing efficiency. Therefore, in such case, allow for extra safety margin when designing your run.

Custom libraries often give 5-10% more percentage of undetermined indexes (with homemade Tn5 for instance) → take that into account when computing required reads/sample.

Spikes cannot be allocated more than 8 mio reads in total.

Versions log

- v1.04: Major edits to the structure to accommodate NovaSeq options. Clarified to use strips only for large numbers of libraries. Added recommendations for qPCR.

- v1.05/v1.06: Indicated that when quantified by qPCR, libraries should be >4nM. Updated "Sequencer and total mio reads requested" fields guidelines.

- v1.07: minor changes (16ul instead of 15ul for novaseq 400mio…).