Sequencing libraries

Sequencing libraries preparation

- NGS is a contamination sensitive technology. When preparing libraries take care to spatially separate pre- and post-PCR steps, storage space and equipment.
- The better quality your RNA/DNA sample has and the larger starting amount (within the protocols specifications) you take, the better the quality and complexity of the library would be.
- Take care not to overamplify the library. Overamplification reduces the library complexity and increases PCR-induced bias.
- If you know in advance which libraries are to be pooled – select proper indices combination BEFORE starting the library prep. Especially for low-plex pooling, consult the Illumina Index Adapters Pooling Guide (link: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjJ5vjHzfvlAhWBLlAKHcgsAwwQFjAAegQIBRAC&url=https%3A%2F%2Fsupport.illumina.com%2Fcontent%2Fdam%2Fillumina-support%2Fdocuments%2Fdocumentation%2Fchemistry_documentation%2Fexperiment-design%2Findex-adapters-pooling-guide-1000000041074-05.pdf&usg=AOvVaw3r0AkWXLx1ah-JCiFG43SO).

Concentration measurement

- Correctly determined **molar concentration** has direct influence on the proper loading of a sequencing flowcell and consequently on the sequencing data output.
- **molar concentration (C_{molar})** can be calculated knowing **mass concentration (C_{mass})** and **average library length (L)**,:

$$C_{molar}\ (nM) = (C_{mass}\ (ng/\mu l) \times 10^6) / (L\ (bp) \times 649\ (g/mol/bp)),$$

  where 649 (g/mol) is an average molecular weight of a base pair.
- To determine **mass concentration** in ng/µl use Qubit and ideally qPCR.
  - Sequencing libraries might contain not only the “correct” adapter-flanked molecules, but also adapter dimers, non-ligated products (especially PCR-free protocols) and non-amplifiable products.
  - Qubit is usually sufficient for RNA-Seq, WGS, exome libraries prepared from >200 ng of starting material and having even, bell-like sizes distribution profiles
  - qPCR (with e.g. KAPA Library Quantification Kit, Roche) is recommended for low-input, PCR-free, low-complexity and specific enrichment libraries. In our experience these are: 10x Genomics, CRISPR -, ATAC-, ChIP-, CROP-, Hi-C- Seq and amplicon libraries.
- To determine **average library length** use BioAnalyser or Tapestation.

- Run your library on BA or TS. Manually assign the library region on the electropherogram, the average length will be determined by the corresponding software.

  NB! this is not just the middle value of the region; the software adjusts for the DNA amounts at all sizes.

- For **molar concentration** calculation:
  - for libraries having even bell-like BA/TS profiles in the 200-700 bp window: use the average length value as determined by BA or TS software;
  - if libraries size profiles have peaks at particular sizes and/or spread above 700 bp – a smaller size value should be used for calculating the molar concentration. Estimate to the best of yours (and ours) experience and correct the value after trial sequencing.
- Less complex and smaller fragments are preferably amplified during bridge amplification on the flowcell, so it is safer to overestimate the molar concentration of the library than to underestimate it.

Contact us if you do not have access to Qubit/qPCR/BioAnalyser or Tapestation

Quality check

- Check that the size distribution profile (BioAnalyser or Tapestation) is as expected (look for examples in the library preparation kit description).
- Narrow peaks at certain sizes usually correspond to overrepresented sequences and reduced library complexity. They might be a library feature (e.g. amplicons). However, if you expect a homogenous fragment lengths distribution, we recommend to redo the library.
- Adapter dimer peaks (around 126 bp for typical Illumina protocols) should be removed by an additional AMPure beads purification.
- Generally, overrepresented sequences and smaller-size library molecules tend to be preferably amplified on the flowcell. A small adapter dimer peak on a library electropherogram, just 5% of the height of the library profile, might result in up to 30% of the adapter reads in the sequencing data.

Libraries dilution and pooling

- Dilute libraries in one of the following diluents (all recommended by Illumina):
  - Tris-HCl 10 mM, pH 8.5 with 0.1% Tween 20
  - Tris-HCl 10 mM, pH 8.5
  - Qiagen Elution Buffer (EB), which is also 10 mM Tris-Cl, pH 8.5
  - Illumina Resuspension Buffer (RSB), which is also 10 mM Tris-Cl, pH 8.5

  We routinely use the buffer with Tween 20. Detergent prevents DNA sticking to the walls of the tube.
- Do NOT combine in one pool:

- complex (RNA-Seq, exome, etc.) and low-complexity (amplicon, CRISPR, etc.) libraries;
- libraries prepared with the same protocol but from samples of different quality (e.g. frozen tissue and FFPE);
- libraries of different sizes.

- For pooling, select proper indices combination. See Illumina Index Adapters Pooling Guide (link: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjJ5vjHzfvlAhWBLlAKHcgsAwwQFjAAegQIBRAC&url=https%3A%2F%2Fsupport.illumina.com%2Fcontent%2Fdam%2Fillumina-support%2Fdocuments%2Fdocumentation%2Fchemistry_documentation%2Fexperiment-design%2Findex-adapters-pooling-guide-1000000041074-05.pdf&usg=AOvVaw3r0AkWXLx1ah-JCiFG43SOBefore

Sequencing libraries amounts required

- If concentration measurement/QC and pooling have to be performed by the Genomics Platforms, provide as much material as possible. Unused DNA and RNA will be properly stored and can be picked up within a year.
- Loading of the library on a flowcell starts with a certain amount of library.
- Input amounts for the available Illumina instruments are summarized in the table below.
- Please provide library stock tubes if library QC (Qubit, BioAnalyser, qPCR) is to be performed at the core facility.

If your samples do not fulfil the standard requirements, please contact us to find an optimal solution.

Table 2. Summary of Illumina platforms minimal library concentration requirements

Instrument	run mode	loading unit	starting loading concentration	loading volume	minimal library amount required	to deliver to the core facility *	
						concentration	volume
iSeq		whole FC	1 nM	5-20 µl depending on the library type	5-20 fmol depending on the library type	≥10nM	5 µl
MiniSeq	HighOutput	whole FC	1 nM	5 µl	5 fmol	≥10nM	5 µl
	MidOutput	whole FC					
NextSeq 500	HighOutput	whole FC	0.5-4 nM	40 - 5 µl	20 fmol	≥10nM	5 µl
	MidOutput	whole FC					

HiSeq 4000		Individual lane	2 nM	5 µl	10 fmol	≥10nM	5 µl
NovaSeq 6000	SP	whole FC	0.5-2.5 nM depending on the library type	100 µl	250 fmol	≥10nM	50 µl
		Individual lane		18 µl	45 fmol		10 µl
	S1	On the FC		100 µl	250 fmol		50 µl
		On each lane		18 µl	45 fmol		10 µl
	S2	On the FC		150 µl	375 fmol		75 µl
		On each lane		22 µl	55 fmol		12 µl
	S4	On the FC		310 µl	775 fmol		150 µl
		On each lane		30 µl	75 fmol		15 µl

* we prefer to have at least the double loading amount of the library or at least 5 µl of 10nM library

Shipment

- Please provide sequencing library (pools) in 1.5ml Safe Lock LoBind tubes, labeled readably with library (pool) name (instructions for naming are in the Library_Transfer_Sheet (link)) and date.
- Sequencing libraries can be sent per post or brought along to our MDC-Mitte or MDC-Buch locations.

  Postal addresses:

  Mitte: MDC-BIMSB, Genomics Platforms, R 1.50, Hannoversche Str. 28, 10115 Berlin
  Buch: MDC, Genomics Platforms, House 64, R 214, Robert-Rössle-Str. 10, 13125 Berlin

  Drop-freezers for self-bringers:

  Mitte: R1.22.1, "Genomics Platforms freezer" – next to the ice machine, Drawer 3
  Buch: House 64, R204, "Genomics Platforms freezer", Drawer 3