LGC Genomics GmbH
Ostendstraße 25, 12459 Berlin, Germany
t +49 (0)30 5304 2200 **f** +49 (0)30 5304 2201
e info.de@lgcgroup.com **w** lgcgroup.com biosearchtech.com

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

Project NGS2468 data delivery, 09 Jun 2021:

Samples:	61
Sequencing amount:	91.5 million read pairs
Sequencing type:	150bp paired-end read (Illumina NextSeq 500/550 v2)
Restriction enzyme(s):	MslI

Delivery contents:

- 'RAW': raw sequencing data after basecalling and demultiplexing in compressed FASTQ format
- 'AdapterClipped': compressed FASTQ files containing sequencing adapter clipped reads
- 'RE_processed': compressed FASTQ files containing restriction enzyme site filtered reads

FastQC v0.11.9 [1] reports, containing read quality metrics, are stored along with the FASTQ files.

Data analysis:

- Demultiplexing of all libraries for each sequencing lane using the Illumina bcl2fastq v2.20 software [2] (folder 'RAW', 'Group' subfolders):
  - 1 or 2 mismatches or Ns were allowed in the barcode read when the barcode distances between all libraries on the lane allowed for it
- Demultiplexing of library groups into samples according to their inline barcodes and verification of restriction site (folder 'RAW', 'Sample' subfolders):
  - no mismatches or Ns were allowed in the inline barcodes, but Ns were allowed in the restriction site
- Clipping of sequencing adapter remnants from all reads (folder 'AdapterClipped'):
  - reads with final length < 20 bases were discarded
- Restriction enzyme site filtering of read 5' ends (folder 'RE_processed'):
  - reads with 5' ends not matching the restriction enzyme site are discarded
- Creation of FastQC reports for all FASTQ files
- Generation of read_counts.xlsx, containing all read counts for all samples at a glance

LGC Genomics GmbH
Bank: HSBC Trinkaus & Burkhardt AG
IBAN: DE94300308800005189004
BIC: TUBDDEDD

UST-ID-Nr.: DE 811547771 | Steuernr.: 37/083/46000
Amtsgericht Berlin-Charlottenburg | HRB 48332 B
Geschäftsführer: Steffi Ulrich | Ross Finlay

LGC

LGC Genomics GmbH
Ostendstraße 25, 12459 Berlin, Germany
t +49 (0)30 5304 2200 **f** +49 (0)30 5304 2201
e info.de@lgcgroup.com **w** lgcgroup.com biosearchtech.com

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

If you have any questions related to your data or some steps of the data analysis, do not hesitate to contact us directly:
Email: bioinformatics@lgcgroup.com
Tel.: +49 (0)30 5304 2202

References

[1] Simon Andrews. *FastQC A Quality Control tool for High Throughput Sequence Data*. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[2] Illumina. *bcl2fastq2 Conversion Software*. URL: https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.