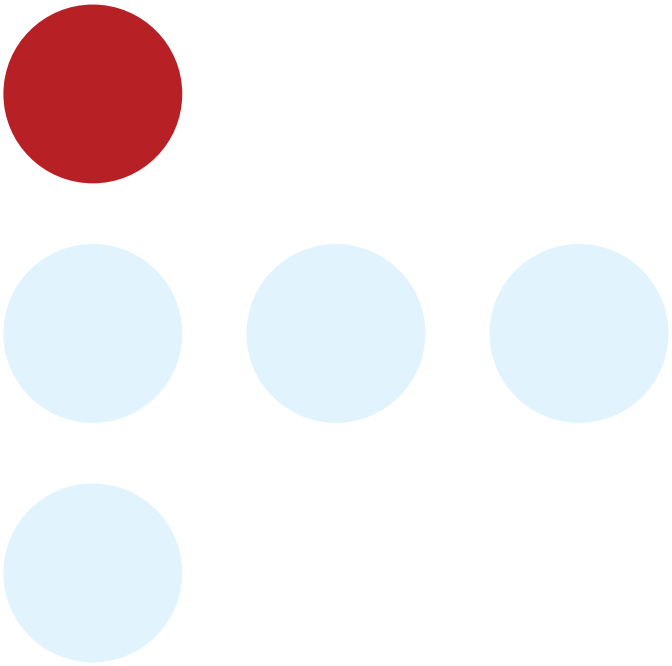
Exiqon
Services

A guide to the microRNA and small RNA Sequencing Service

Guidelines v2.2
September 2017



EXIQON
Seek Find Verify

Table of Contents

Consultation and experimental design . . . 3
How do I get started? . . . 3
Designing the optimal experiment . . . 3
microRNA sequencing or small RNA sequencing? . . . 3
Effect of sequence reads on new microRNAs discovery . . . 4

RNA sample submission . . . 6
How do I isolate RNA? . . . 6
How much purified RNA is needed? . . . 7
If you want Exiqon to isolate the total RNA . . . 7
How do I assess the RNA quality? . . . 8
How does Exiqon assess the RNA quality? . . . 9
RNA integrity . . . 10
qPCR-based RNA sample quality control (for serum/plasma NGS only) . . . 10
Library preparation and QC . . . 10
Sequencing . . . 11
How many biologial replicates do I need? . . . 11
How do I send my samples? . . . 11
What about customs regulations? . . . 12
Receipt of samples at Exiqon . . . 12

Data analysis . . . 13
QC and mapping . . . 14
Normalization . . . 14
Differential analysis . . . 14
Biological interpretation of results . . . 14

Report and final consultation . . . 15
Security Statement . . . 15

Frequently asked questions . . . 16

Wordlist . . . 19

References . . . 20

Consultation and experimental design

How do I get started?

Please contact Exiqon to discuss your planned project and arrange a free consultation with our NGS experts.

- If you decide to use Exiqon's Next generation Sequencing (NGS) Service, experimental details should be submitted using the Sample Submission Form (SSF) available at www.exiqon.com/ssf, in order for Exiqon to obtain all the information necessary to perform the experiment.
- Upon submission of the SSF, a copy is sent to Exiqon where it will be reviewed and approved by an Exiqon scientist. You will be given a unique project reference code. Please print out and sign the approved form. Enclose the signed form with the samples upon shipping.

Designing the optimal experiment

- When you engage Exiqon for your small/microRNA NGS Service projects, you are assured direct communication with the scientists performing your experiments throughout the duration of the project.
- Each project begins with a free consultation with an RNA NGS expert. Together we design an experimental setup that best satisfies your research needs and budget. By referencing the completed and detailed sample-submission form we can ensure that all experimental details and subsequent analysis are clearly defined and understood by both parties.

microRNA sequencing or small RNA sequencing?

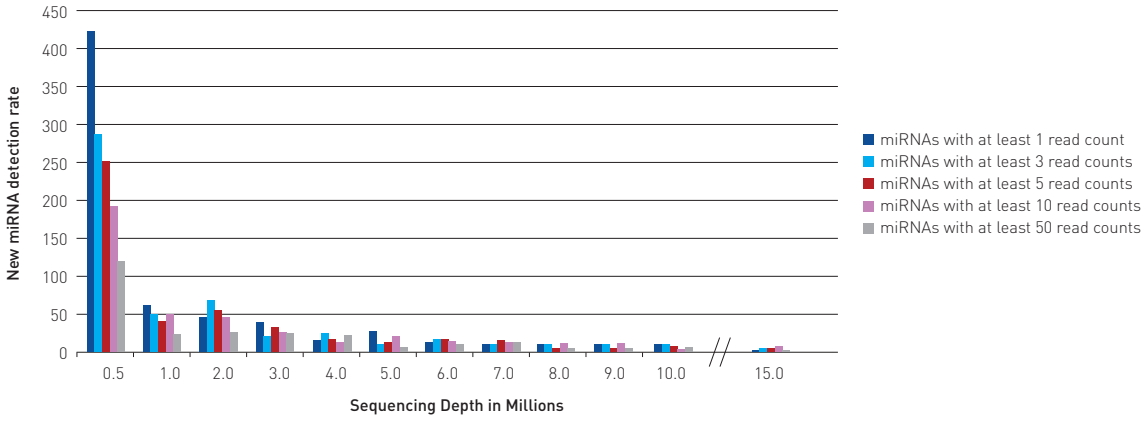
- Exiqon offers two small RNA sequencing products; microRNA sequencing and small RNA sequencing. MicroRNA sequencing is specifically focused on discovery and quantification of microRNAs. This includes novel microRNA detection and prediction. Small RNA sequencing is specifically for other small RNA targets such as sn/sno/tRNA.
- We recommend microRNA sequencing to those doing microRNA research.
- MicroRNA sequencing is dependent on automated gel extraction of a band representing insert size of 15-30 nt. in length; this assures that contamination of degraded RNA, empty adaptors or primer dimers is minimal in the NGS microRNA library.
- Small RNA sequencing is dependent on excision of a custom size for library construction (within 30-200 nt. range), which can include microRNAs, pre-microRNAs, tRNAs and snRNAs, snoRNAs and other small RNA molecules in this size range. We recommend choosing a biological relevant size band within this range to get as large a proportion of relevant reads as possible.
- Many small RNAs have protective caps (for example reversed orientation 7-methyl- guanosine) on their 5' end or modified bases (for example 2' O-methyl groups) on the 3' end, making ligation of adaptors highly challenging. We take steps to maximize the success of the adaptor ligation steps, but there is no guarantee that modified or capped RNA species will be present in the libraries. We therefore recommend small RNA sequencing only when the targets of interest are small RNAs other than mature microRNAs.
- We recommend small RNA sequencing to those researching in snRNAs, snoRNAs, tRNAs or other small RNAs under 200 nt in size. Please note that we may not be able to compare expressions levels of RNA species with different lengths.

- Small RNA sequencing can only be performed on high quality RNA. Partially degraded RNA or RNA from FFPE samples contains large quantities of degraded rRNAs resulting in a large portion of the sequencing reads being irrelevant to any discovery or analysis. Thus, it may be necessary to sequence with a higher number of reads to capture a meaningful fraction of relevant reads.

Effect of sequence reads on new microRNAs discovery

The microRNA library read depth is one of the most crucial factors with regards to both differential expression analysis and discovery of novel microRNAs. Metpally et al. analyzed Next Generation Sequencing microRNA data from Gastrocnemius muscle in mice and corresponding plasma samples using different read number from 0.5-15 million raw unmapped reads. They observed that read numbers greater than 5 million contributes very little to the detection of new microRNAs (Metpally et al. 2012 Front Genet. 4(20)).

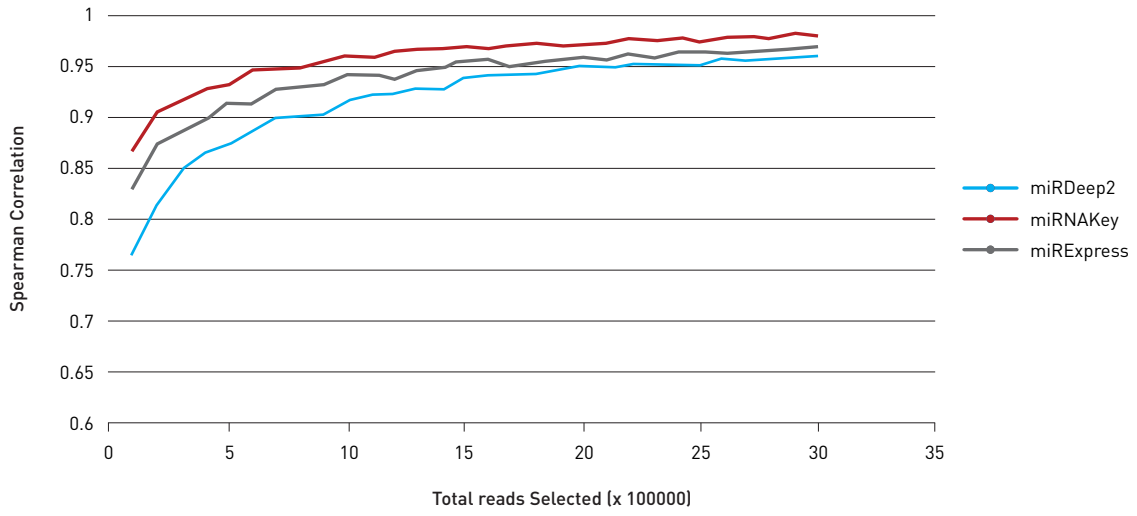
Figure 1. Effect of read depth on discovery rate of rare or novel microRNAs. As 1 million reads (filtered but unmapped) at a time are added to the sequencing depth (from a gastrocnemius muscle sample), the number of additional novel microRNAs that are detected is reduced. Each bar on the chart shows the numbers of newly detected microRNAs with at least 1, 3, 5, 10, or 50 reads respectively at different depth (Metpally *et al*. 2012).



Metpally et al. also investigated the correlation between different mapped read numbers from 100,000 to 3 million reads using three different analysis tools (miRDeep2, miRNAKey and miRExpress). They found that the correlation slowly increases but only of about 0.05 from 1 to 3 million reads. These numbers can vary from tissue to tissue but at Exiqon we are confident that 7.5 million reads are sufficient for comprehensive discovery and powerful differential expression analysis.

When performing small RNA sequencing the reads will contain more of snRNAs, snoRNA, tRNAs and other small RNAs why more sequencing reads may be needed for comprehensive analysis. If required, it is possible to add 7.5 million reads for small RNA sequencing at a small additional cost.

Figure 2. The correlation of a subset of reads, beginning with 100,000 with the full number of mapped reads, 3.5 million. The correlation becomes fairly stable at 1.5 million random reads, correlation coefficient of 0.97for miRNAKey. Our main observation was that increasing the mapped reads from 1 to 3 million reads only increased the Spearman correlation by 0.05 (Metpally et al 2012).



RNA sample submission

How do I isolate RNA?

High quality samples are important for accurate sequencing. During the initial consultation, we offer recommendations on suitable extraction and clean-up methods. Some of our recommendations on isolation of RNA are listed below. Exiqon also offers RNA isolation in addition to its profiling service. (Details on Exiqon's isolation services can be found below in section "If you want Exiqon to isolate the total RNA", page 7).

- Total RNA should be prepared using a method that preserves small RNA species. When using commercially available kits, please note that they often have two alternative protocols. One is for isolation of total RNA (including microRNA) and one is for enrichment of small RNA. For both microRNA and small RNA sequencing we recommend using the protocol for total RNA if it also retains the small RNA species.
- The purified RNA should be eluted or dissolved in RNAse-free water.
- No carriers or spike-ins should be used in the purification protocol.
- Avoid heparin in collection tubes and cell culture media as it can inhibit downstream enzymatic reactions resulting in sub-optimal library generation.

Purified microRNA

- We recommend total RNA as starting material, but high quality purified microRNA is also accepted

Formalin-fixed, paraffin-embedded (FFPE) tissue

We recommend only using FFPE RNA isolation kits that also preserve small RNAs. FFPE derived RNA can only be used for microRNA sequencing with strict size selection. Small RNA sequencing on FFPE derived material is not possible. Note that after the library preparation, a significant portion of the sequencing reads may be from degraded RNA species (rRNA/mRNA/tRNA) and therefore extra read depth might be needed on FFPE material.

Tissue and cells

We recommend the Exiqon miRCURY™ RNA Isolation kits **www.exiqon.com/rna-isolation** purification of RNA from tissue or cells.

Note: RNA isolation aimed at NGS profiling should not be isolated with RNA carriers or spike-ins as these will interfere with the library generation.

Whole blood: We recommend using a commercial kit designed for purification of total RNA from whole blood. The kit of choice may depend on the type of tube used for collection of the whole blood. During your initial consultation Exiqon will be happy to advise on the most appropriate RNA isolation method for your samples.

How much purified RNA is needed?

The table below (table 1) shows how much total RNA we recommend to send for microRNA or small RNA sequencing. It is recommended to send material for 2 library preparations in case yields are low from the library preparation and/or a re-run is needed. If you are unable to provide the minimum recommended amounts, please contact us to discuss sequencing using lower input amounts of RNA.

We recommend that you send total RNA with a minimum concentration of 25 ng/µl.

Table 1. Amounts of total RNA and FFPE needed for microRNA or small RNA QC and sequencing.

Sample type	Minimum amount needed for QC (NanoDrop/bioanalyzer)	Minimum amount needed per library preparation	Total amounts of total RNA recommended
Total RNA	60 ng	100 ng	300 ng
FFPE*	60 ng	1 ug	2 ug

*If FFPE only Nanodrop QC is performed

If you want Exiqon to isolate the total RNA

Exiqon offers an RNA isolation service in addition to its profiling services. Due to personal health safety legislation, we do not accept any form of contagious material, or samples shipped in glass tubes. All samples must be shipped to Exiqon Services in clearly labeled 1.5-2.0 mL plastic tubes.

Blood serum and plasma
All samples should preferably be collected using the same protocol to minimize technical variation. Examples of good protocols for collection of samples can be found at:

Plasma SOP:
http://edrn.nci.nih.gov/resources/standard-operating-procedures/standard-operating-procedures/plasma-sop.pdf

Serum SOP:
http://edrn.nci.nih.gov/resources/standard-operating-procedures/standard-operating-procedures/serum-sop.pdf

Upon collection of plasma or serum - prior to freezing - spin at 3000 g for 5-10 min to collect debris. Take the supernatant from the vial and leave some plasma/serum on top of debris. This reduces the potential contamination of cellular components (from e.g. buffy coat) even further.

If you want Exiqon to isolate the total RNA, please see the chart below for amounts needed:

Table 2. Amounts of sample needed for purification and subsequent microRNA or small RNA NGS Services.

Sample type	Amount needed	Comment
Plasma/ serum (human)	500 µl	microRNA-seq only
FFPE sections	Min 6 x 10 µm sections of 1 cm^2	Not mounted on slides.
Fresh-frozen tissue	4-5 mg	Larger amounts of tissue may be required for tissues with low RNA content e.g. bladder, bone, adipose.
Cells	2 x 10(6) cells pelleted and frozen	Spin cells down gently, take the medium off, rinse gently in cold PBS once, remove the PBS, and freeze quickly (e.g. liquid nitrogen) and store at -80 degrees Celsius.

Please consult us for instruction on how to collect samples for purification prior to shipping samples to Exiqon Services.

How do I assess the RNA quality?

Exiqon will perform RNA quality control prior to profiling, but we recommend that you check your RNA yourself prior to submission. We recommend measuring the OD260/230 ratio as well as the OD260/280 ratio. See details below for the rationale behind these measurements. If any of these ratios are lower than 1.6, it may be advisable to perform additional column purification that preserves small RNA in order to be absolutely sure of superior sample quality.

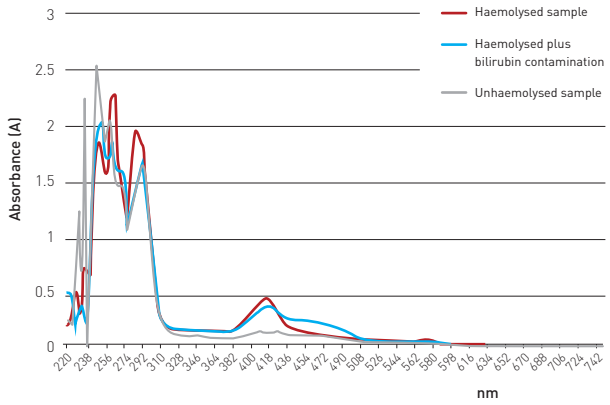
If possible, we also recommend checking the integrity of total RNA using a BioAnalyzer RNA assay prior to shipment, to avoid increased turn-around-time (TAT) due to re-submission of samples.

Small RNA or microRNA sequencing is best performed on high quality RNA. Partially degraded RNA or RNA from FFPE samples contain large quantities of degraded rRNAs resulting in a large portion of the sequencing reads being irrelevant to any discovery or analysis. It is also very important the RIN values are not associated with grouping of the samples in a given experiment. Please contact us if you are interested in sequencing FFPE samples or samples with RIN value below 7.

For serum or plasma samples
It can be advantageous to perform a hemolysis check on the samples prior to shipment. To monitor hemolysis we recommend two steps:

1. Visual inspection. Pink or red samples are hemolysed. White or yellow samples are usually not hemolysed. But the range from yellow to pink can be difficult to interpret.
2. Spectrophotometry. By measuring the absorbance in the range from 230-600 nm of the undiluted plasma/serum sample, the absorbance of hemoglobin can be analyzed. If there is a distinct peak at 414 nm, then there is oxyhaemoglobin in the sample indicating hemolysis. Examples of hemolysed and non-hemolysed samples are shown below (figure 3).

Figure 3. Examples of three spectra obtained from different plasma samples. An un-hemolysed sample (gray), a hemolysed sample (red) and a hemolysed sample with a shoulder (a bilirubin contamination (blue)).



How does Exiqon assess the RNA quality?

When RNA samples arrive at Exiqon they undergo a quality assessment prior to the NGS profiling analysis. The standard quality assessment includes absorbance measurements and RNA integrity measurements.

Absorbance spectrum

To assess the purity of the samples, we examine the absorbance spectra to identify potential contaminations and differences between the samples in the same project. Differences in purity, or obvious contaminations, may affect the downstream results. For most projects these contaminations do not have an effect, but they could impact projects in which very minute biological differences are being investigated.

- OD260/230 nm <1.6: indicates potential contamination with Guanidinium isothiocyanate (GdSCN) or other chaotropic agent absorbing at 230 nm. This is seen if the wash buffer is carried through in column purifications.
- OD260/ 280 nm < 1.6: indicates potential contamination with phenol absorbing at 270 nm. This is seen if part of the phenol phase is aspirated when collecting the aqueous phase in a phenol:chloroform extraction.

Both contaminations may reduce the performance of the library preparation.

RNA integrity

To assess the integrity of the ribosomal subunits, a Bioanalyzer profile is run and an RNA integrity number, RIN, value is determined by the software. The RIN value is a measurement of the intactness of the two ribosomal bands. RIN value >7: High quality RNA. Please note that for samples with RNA concentrations below 25 ng/µL, the robustness of RIN values is poorer. Best results are obtained for concentration values above 50 ng/µL.

For FFPE samples, we do not measure RIN values. Only microRNA NGS is recommended for FFPE derived RNA. Sequencing FFPE derived microRNA libraries will result in a significant portion of the reads to be degraded RNA species (rRNA/mRNA/tRNA).

In general, for sequencing:
- RIN 7-10: Little or no degradation; microRNA or small RNA sequencing possible
- RIN value 5-7: Partial degradation; microRNA sequencing possible
- RIN value <5: Degraded RNA; microRNA sequencing possible but sequencing of degraded material unavoidable

Recommended
1 - High quality RNA
2 – RNA of similar quality (e.g. all degraded)

Avoid
1 – Large differences in RIN values between samples
2 – One biological group with high quality RNA, one with degraded RNA.

qPCR-based RNA sample quality control (for serum/plasma NGS only)

In Exiqon services, we have developed a unique qPCR based QC-panel to test RNA quality from serum and plasma samples prior to library preparation.
- We do not recommend pooling of samples, neither from sample preparation or purified RNA. This usually does not give valuable information about the samples
- The microRNAs to be tested are pre-selected by Exiqon
- The QC screen is based on running all samples using standard experimental setup for the miRCURY LNA™ Universal RT microRNA PCR system. Six pre-selected microRNAs and four synthetic spike-ins are assayed. Positive and negative control samples are included.
- The levels of microRNAs will be evaluated across all samples based on experience and compared to prior data from similar sample sources. For plasma and serum samples, the degree of hemolysis will be evaluated.
- This RNA QC allows removal of outlier samples from the sample cohort, but optimizing the experimental setup is not possible.

Library preparation and QC

In Exiqon Services, we perform two types of sequencing library quality controls. After libraries are generated using 3' & 5' RACE like protocols, Reverse Transcription (RT) and PCR pre- amplification, the insert rate of the desired RNA type is evaluated using a Bioanalyzer DNA high sensitivity chip. If the library is deemed to have sufficient inserts for downstream processing it is size fractionated using a Caliper XT gel cutter and the desired band(s) are extracted. Then qPCR based quantification of each library is performed, and samples are normalized and pooled in equimolar ratios. After pooling of sample libraries, qPCR based quantification is performed on the library pool to ensure optimal concentration for cluster generation on the flowcell.

Sequencing

The library pool(s) to be sequenced are denatured and diluted/neutralized in the required concentrations. Then cluster generation is performed on the apopriate folwcell using single molecule clonal amplification. Finally, the high-throughput next generation sequencing is performed using the Illumina sequencing technology platform. For more detailed description of the sequencing process, please visit the Illumina homepage at www.illumina.com

How many biological replicates do I need?

- The number of biological replicas needed for microRNA and small RNA sequencing depends on the objectives of the experiments. Inclusion of at least three biological replicates per sample group will allow statistical tests of data comparisons, but we recommend a minimum of four.
- Overall the reproducibility of technical replicates is such that we recommend prioritizing biological replicates over technical replicates for most screening purposes.

How do I send my samples?

- Please ensure that you include a signed copy of the sample submission form with your samples.
- Ensure that samples are labeled clearly and with unique numbers using a permanent marker. Pack the samples arranged in the same order as listed on the SSF in a cryo storage box or similar.
- Please use the fastest available shipping service. For international shipping please use a courier service such as FedEx.
- If you are shipping from outside Europe and North America, please only send your samples on Monday or Tuesday to avoid weekend deliveries.
- RNA samples should be shipped on dry ice in Styrofoam insulated boxes. Please make sure you use adequate amounts of dry ice. We recommend a minimum of 3.5 kg for shipments in North America, 9 kg for shipments from Asia and Australia and 6 kg for Rest-of-the-world.

For North America, Mexico and Canada, please ship samples to
QIAGEN Genomic Services
Attn. Krishna Amin
QIAGEN Americas
6951 Executive Way
Frederick, Maryland 21703
USA
Phone: +1 301 673 5045

For all other countries, please ship samples to:
QIAGEN GmbH
R&D Life Science Key Account Service
Attn. Andre Bahr / Anke Singer / Holger Wedler
Qiagen Str. 1
40724 Hilden
Germany
Phone +49 2103 29 11649

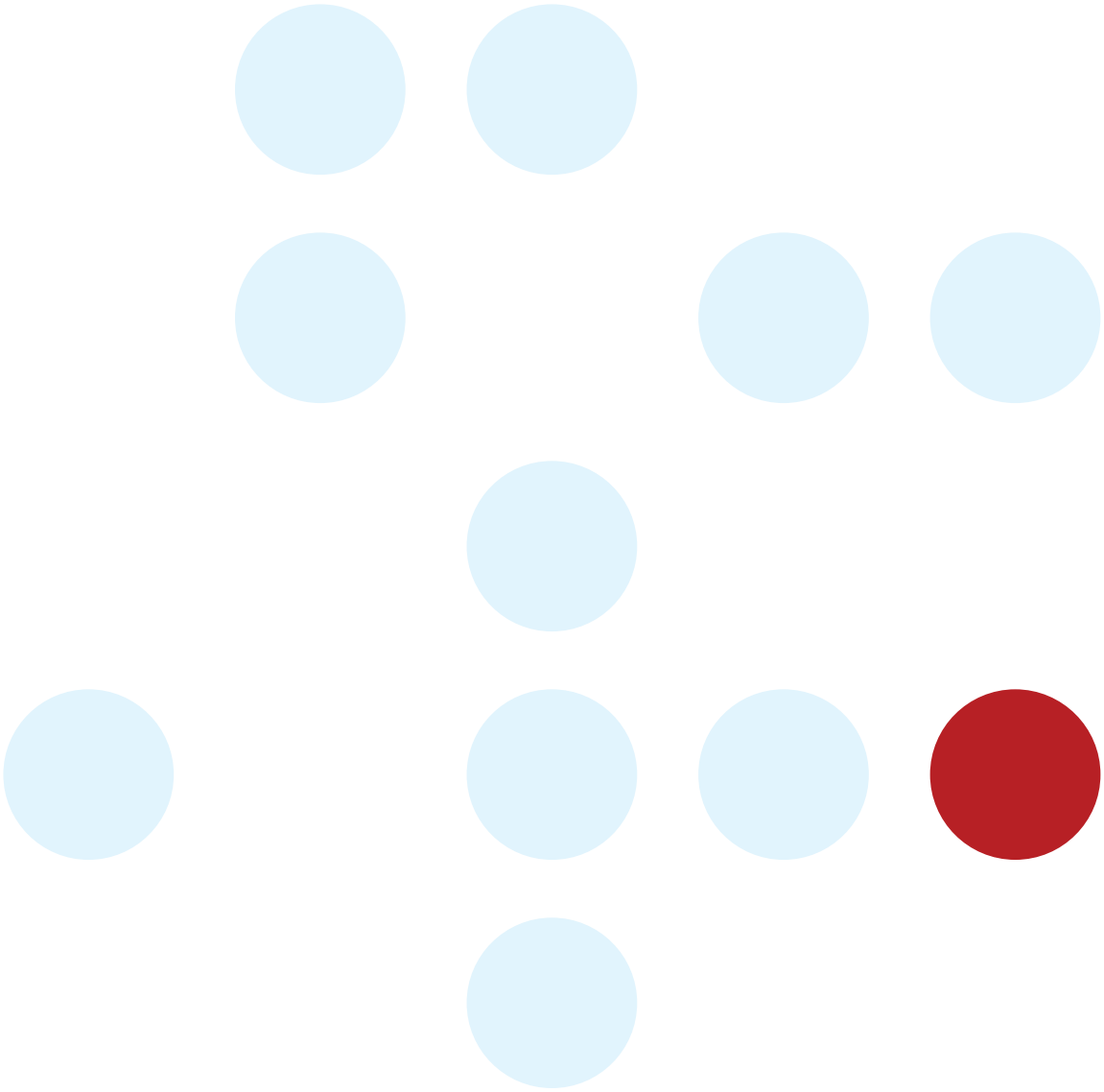
What about customs regulations?

In order to avoid delays in customs, please make sure that you describe the content accurately: i.e. Purified RNA dissolved in water, for research purpose only.

- The commercial invoice should state a value of 0 (or lowest amount possible)
- For non-human samples, please contact your local representative

Receipt of samples

Upon receipt of your samples we immediately transfer them to a secure –80 degrees Celsius freezer. You will receive an e-mail confirming that your samples have arrived in good condition.



Data analysis

The data analysis represents a crucial step in the NGS pipeline, not only to perform a biological interpretation of the data, but also to evaluate the quality of the data and the sample. As part of our NGS service we provide a comprehensive data analysis appropriate for specific experiments and individual needs.

The bioinformatics is an integrated part of our NGS platform and our scientists have a strong background in both the experimental and analytical aspects of Next Generation Sequencing. This means that, rather than applying a standard analysis pipeline to all projects, we consider each project independently to determine the most appropriate analysis to answer the relevant biological questions. Our data analysis typically involves data QC, including characterization of the read populations, differential expression analysis, and comparison of sample sets (unsupervised and supervised clustering via two dimensional heat maps, PCA plots, and pairwise comparison and visualization by volcano plots). Specific details are given in the following sections.

Table 3. List of species currently supported (for other species, please inquire). Please note: Serum/plasma NGS only human is supported.

Species	Common name	Three letter nomenclaure miRBase
Homo sapiens	human	hsa
Mus musculus	mouse	mmu
Rattus norvegicus	rat	rno
Arabidopsis thaliana	rockcress	ath
Bos taurus	cow	bta
Caenorhabditis elegans	roundworm	cel
Canis familiaris	dog	cfa
Danio rerio	zebrafish	dre
Drosophila melanogaster	common fruit fly	dme
Gallus gallus	chicken	gga
Glycine max	soy bean	gma
Macaca mulatta	rhesus monkey	mml
Oryza sativa japonica	rice	rno
Pan troglodytes	chimpanzee	ptr
Sorghum bicolor	durra	sbi
Sus scrofa	pig	ssc
Zea mays	corn	zma

QC and mapping

As a first step, the read data is subjected to a rigorous QC step to investigate the fidelity of the data, removal of low quality reads and evaluation of the content of the reads. To do this, the reads are compared to a number of reference sources.

In particular:

- Adapters are trimmed and subsequent reads representing high quality sequencing data are mapped to a reference genome (species at standard price are listed in table 3) annotated with miRBase in order to evaluate mapping percentage and identify reads associated with microRNAs.
- Unmapped reads are queried against additional reference sources (such as Rfam) to identify additional classes of small RNAs, potential novel microRNAs and sources of contamination that might be present in the sample.
- A table of known and putative microRNAs and their associated number of reads is then generated for differential expression analysis. This forms the micro- or small RNA- sequence profile for the sample.

Normalization

- The differential expression analysis investigates the relative change in expression (i.e. reads) between different samples. Thus, the normalization is primarily concerned with compensating for sample specific effects (generally caused by the variation in sequencing depth between samples). Additionally, the normalization step offsets under-sampling effects (due to highly expressed microRNAs dominating the read set) by identifying scaling factors that minimize log fold changes between samples across the majority of microRNAs. These scaling factors are then applied to the counts to generate a corrected count table.

Note: This corrected table is only used for the differential expression analysis and the count table presented in the report represents the true number of counts for each microRNA.

Differential expression analysis

- The differential expression analysis step attempts to distinguish biological variation from technical variation within the experiment based on the assumption that this varies amongst microRNAs. For single factor experiments we use the quantile-adjusted conditional maximum likelihood method as studies have shown it to be most effective for NGS experiments (Robinson and Smyth, 2007). For more general experiments containing multiple factors, we use the Cox-Reid profile adjusted likelihood method. A list of microRNAs predicted to be differentially expressed between two experimental conditions is presented in the final report.

Biological interpretation of results

To help with understanding the significance of the results, we perform a final step where we provide a biological interpretation of the findings. This can be particularly useful when novel or less common microRNAs are predicted to be differentially expressed in the sample.

- First of all, we perform a Gene Ontology (GO) enrichment analysis to determine which GO terms are overrepresented in the differentially expressed microRNAs. We obtain the GO terms via the annotation associated with the microRNA targets. This information is linked to our miRSearch database, a rich source of curated information related to microRNAs and their targets.
- Secondly, we further use the freely available miRSearch database to provide additional data associated with the differentially expressed microRNA set and aid with the downstream experiments. For each microRNA this information includes experimentally verified target information, primer design, and other relevant information for the particular microRNA.

Report and final consultation

The data is presented in a PDF summary report containing a description of the project, assessments of sample - and data quality and an overview of the results analysis with publication- grade illustrations. Extensive Excel files with all the major findings and statistical analysis is provided. On request we supply all FASTAQ files, mapping files (BAM) on an encrypted USB stick (additional charges apply).

A microRNA Next Generation Sequencing Service sample report can be downloaded from our webpage **www.exiqon.com/ngs-services**.

- Upon completion of your NGS project, we will send you an e-mail with a link to a secure web-server from which you may download the final report and associated spreadsheet files. Exiqon offers a free consultation with one of our microRNA experts to discuss the data and how to best set-up validation experiments.
- For standard projects please allow approximately 4-6 weeks for completion of the entire analysis from samples passing RNA-QC to delivery of the final report. For larger projects (> 40 samples) longer turn-around times are to be expected. We will keep you informed about the progress of your profiling experiments throughout the course of the project. Furthermore, you will have instant access to project status via your personal login on exiqon.com
- Exiqon will ensure that all information are stored confidentially and will not use any data generated for purposes other than customer statistics and internal quality control.
- Exiqon will store all raw data for 3 months on secure servers after delivery of the final report. On request we supply all FASTAQ files, mapping files (BAM) on an encrypted USB stick.
- The unused portion of any samples will be discarded 3 months after completion of the experiments unless specifically agreed otherwise.

Security Statement

- IT security is very important to us at Exiqon, and as a result, all security standards are derived from ISO 17799/DS484 and BS 7799 standards. We strive to deliver a very high level of data security and reliability, and enforce yearly IT audits to make sure that IT policies and security standards are in place and being complied to.
- To ensure customer confidentiality and integrity, Exiqon has chosen to maintain a local and centralized infrastructure platform, complete with quick failover points and redundant hardware. This is maintained by our own IT-department in collaboration with IT experts.
- Exiqon´s main data center meets the Tier level 4 data center requirements, complete with alarm based access restriction and environment sensoring.
- Our server infrastructure is approx. 95% virtualized, running on a uniform and redundant equipment platform. All servers/storages are using RAID configurations, (Raid 5, 6 or 10) depending on application requirements.
- Management of employee access rights are reviewed by department heads according to their job functions. These access rights are then delegated, by trusted IT employees in accordance with Microsoft standards (AD and GPO).
- Exiqon has a comprehensive and redundant backup and disaster recovery strategy. Local data backups are made daily and are kept in a secondary tier level 3 datacenter. All backups are then replicated off-site to a trusted business partner, in compliance with our disaster recovery plan.
- Our external security is based on a dual UNIX/Linux Firewall platform, and all customer related web portals are placed in secure DMZ-zones.

Frequently asked questions

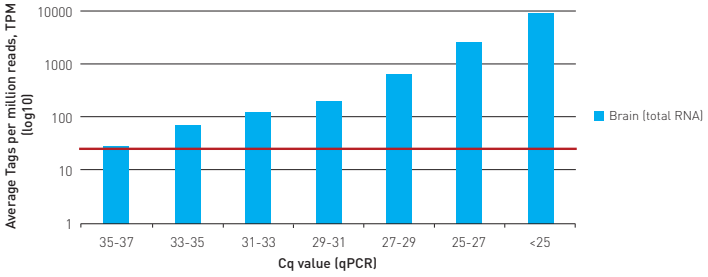
Question: Are 7.5 million reads enough for discovery and differential expression analysis?
Answer: Yes they are. Metpally et al. has shown that approximately 5 million reads will provide you with the statistical power necessary to analyze your data and achieve a high discovery rate of novel microRNAs (Metpally et al., 2012).

Question: I am interested in validating several novel and known microRNAs. What do I do?
Answer: Exiqon offers validation of NGS results on Pick-&-Mix qPCR plates using our extensive pre-designed and validated microRNA assays. We also offer custom designed qPCR assays for novel putative microRNAs. However, it is important to consider the read counts of the microRNAs of interest. Internal microRNA NGS data has shown that microRNAs with low read count can be hard to validate (see figure 4).

We use Cq values of 37 as a cutoff to ensure robust and reproducible data points. Consider this when selecting assays for validation. Note however these are averages and affected by tissue type and the microRNA read number as well as microRNA numbers and distribution within a given sample. Exiqon Service Scientists can help you design the optimal experimental setup to maximize your likelihood of success.

Question: Can you analyze FFPE samples? Do the chemical modifications interfere with library preparation? Do we need higher number of reads per sample as many reads will be wasted on degraded RNA fragments?
Answer: Yes microRNA sequencing is possible on FFPE material. Sequencing FFPE derived libraries will result in a significant portion of the reads originating from being degraded rRNA species. Please contact us if you are interested in sequencing FFPE samples.

Figure 4. Comparison of NGS data and miRCURY LNA™ Universal RT microRNA PCR data (Human panel I+II). microRNAs with counts above the red line have good chance of being validated using Exiqon's qPCR platform. At low counts it is very difficult to get consistent signal on a qPCR platform. However, microRNAs with approximately 25 tags per million reads (TPM) or above will most likely give a robust signal upon qPCR validation. 25 TPM corresponds to Cq values of approximately 35. Note however these are averages and affected by tissue type and the microRNA number and distribution within the sample.



Question: Which data points were removed due to data quality controls?
Answer: Data for analysis needs to pass the following criteria:
- High quality score (Q score) and read length of >15.
- Be mappable to the corresponding genome(s) and/or databases.
- Pass background filtering based on read numbers (to remove low copy reads).

If the data points do not meet these criteria, they are removed from the dataset.

Question: What is multiple testing corrections?
Answer: When a large number of statistical tests are carried out simultaneously, ordinary p-values need to be adjusted in order to control the number of false positives, i.e. the number of genes for which the null hypothesis 'the gene is equally expressed between groups' is incorrectly rejected – type I errors (Benjamini and Hochberg, 1995).

Question: What is GO analysis?
Answer: The Gene Ontology is a formal representation of species independent knowledge associated with genes and their products. This means the information can be parsed and analyzed by computers to associate the knowledge with the results from biological experiments in order to gain further insight.

Question: What is the average read length for microRNA sequencing and for small RNA sequencing?
Answer: The read length is defined in the run settings on the machine. For microRNA and small RNA we use 50 cycles (read length of 50 nt). For other settings, please contact us and we will set-up a custom project for you.

Question: Will you be doing barcoding and multiplexing samples together?
Answer: All NGS libraries will be made individually and can therefore be compared individually as well. However the samples are pre-amplified with different Index primer sets that can be pooled together for appropriate sequencing pools after library quantifications and QC.

Question: What adaptors do you use?
Answer: We use the adaptors that are supplied with the NEBNext® Small RNA Library Prep kit (New England Biolabs). Exiqon will select which indices that will be used depending on availability and suitability.

Question: How does Exiqon deal with isomiRs in the analysis?
Answer: We examine the reads mapped to specific microRNAs to identify the presence of isomiRs. These can be distinguished or summed together in the final report according to your requirements. It is important to state in the SSF if isomiR analysis is required.

Question: Which normalization procedures are needed for NGS data? E.g. to account for barcode-induced variation, library size, or just normalization based on mean read counts?
Answer: For NGS data, normalization is primarily concerned with compensating for sample specific effects generally caused by the variation in sequencing depth between samples. The normalization step also offsets under-sampling effects (due to highly expressed microRNAs dominating the read set) by identifying scaling factors that minimize log fold changes between samples across the majority of microRNAs.

Question: Will Exiqon start with total RNA or enriched small RNA? Or will Exiqon size- fractionate the RNA?
Answer: We recommend total RNA, but can make libraries from both RNA preparations. We do the size selection in the library preparation step.

Question: What will be the minimum RNA amount needed?
Answer: It depends on the application and sample type. Please refer to table 1, page 7 for recommended input amounts.

Question: How will Exiqon remove adaptor dimers?
Answer: After the library preparation we use an automated gel cutter. By knowing the expected sizes of both the adaptor dimers and the library it is possible to select only the fraction containing library, excluding the primer dimer fraction.

Question: Will barcodes be introduced during the PCR or in the adaptors?
Answer: The indices are added during the PCR where they are part of the 3'-primer.

Question: How will Exiqon select the microRNA size range?
Answer: We use an automated gel cutter (Perkin Elmer Labchip XT). This allows for selection of molecules in the microRNA size range.

Question: What data filtering will be needed?
Answer: We trim adapters and then filter by length to remove short reads (< 15 nt). Reads are then mapped to a series of reference sequences (reference genome, miRBase, other RNAs and potential library contaminants). We have specific reference sets but any of these steps can be tailored to specific customer needs. For example, we can use different cut off length or apply a different version of miRBase.

Question: What quality (RIN number) should the RNA have? Degraded RNA fragments could enter the library, so do we need higher quality criteria than normal?
Answer: The higher your RIN values are, the less affected the data will be by degraded RNA leading to higher quality results. We recommend RIN of > 7 although lower RIN values are accepted for microRNA sequencing. However note that lower RIN values will result in a greater portion of the data coming from degraded ribosomal RNA which will have no meaning for the discovery or analysis.

Question: Will 2'OMe modified piRNAs be included in the library?
Answer: Yes, if piRNAs are present in the sample, we can include this RNA species in the library and sequencing. Our sample preparation kit can ligate the adaptors onto this modification, however the efficiency compared to unmodified nucleotides is impossible to evaluate since it is also dependent on the 3' sequence composition. Please contact us if you are interested in piRNA sequencing.

Question: Is it possible to expand the groups later?
Answer: Yes, it is possible to add more samples to groups at a later date. Additional bioinformatics analyses will be requested as custom service.

Wordlist

FASTQ	Format for storing sequence data along with quality scores from a sequencing run. The quality format uses a single ASCII character. Can be opened in standard text editors. Big files can be difficult to handle by the editor.
BAM	Binary Alignment/Map format. Can be accessed through BEDtools or similar softwares.
Mapping	Process of finding the location of each read in a genome using a reference database.
Q-score	Or a Phred-score. A score given to each base describing the quality of the base call at the specific position. The quality score is logarithmically related to the base-calling error probabilities. The scale goes from 1 to 40. A Phred score of 30 equals a base call accuracy of 99.9 % and is accepted as cut-off for high quality data. A phred score of 20 equals a base call accuracy of 99% and is considered medium quality data.

References

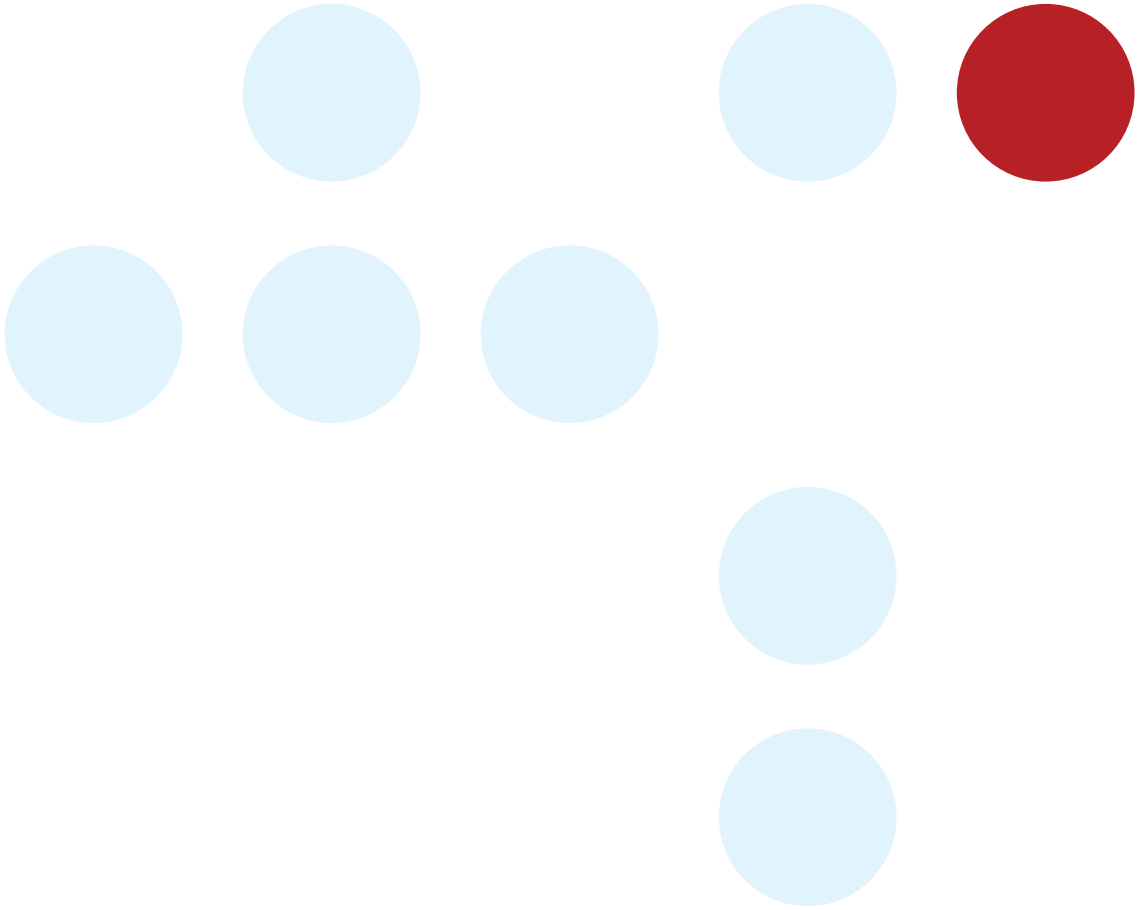
Benjamini and Hochberg (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B, 57, 289-300.

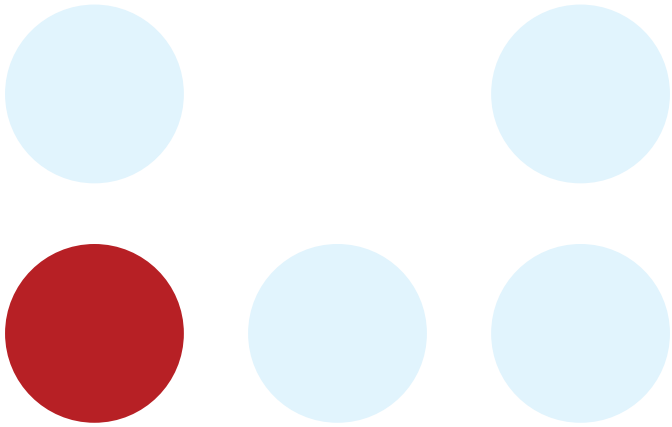
Buermans, H.P.J. *et al.* (2010) New methods for next generation sequencing based microRNA expression profiling. BMC Genomics, 11, 716.

Metpally R.P.R. *et al.* (2012) Comparison of Analysis Tools for miRNA High Throughput Sequencing Using Nerve Crush as a Model. Front Genet. 4(20), 1-13.

Robinson, M.D. and Smyth, G.K. (2007): Moderated statistical tests for assessing differences in tag abundance. Bioinformatics, 23(21), 2881-2887.

Wang, Y. *et al.* (2011) Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. BMC Bioinformatics Proceedings of the Eighth Annual MCBIOS Conference. Computational Biology and Bioinformatics for a New Decade, College Station, TX, USA. 1-2 April 2011.





Outside North America
QIAGEN GmbH
R&D Life Science Key Account Service
Attn. Andre Bahr / Anke Singer / Holger Wedler
Qiagen Str. 1
40724 Hilden
Germany
Phone +49 2103 29 11649

North America
QIAGEN Genomic Services
Attn. Krishna Amin
QIAGEN Americas
6951 Executive Way
Frederick, Maryland 21703
USA
Phone: +1 301 673 5045

www.exiqon.com

EXIQON
Seek Find Verify

923007 - v2.2 - 09/2017