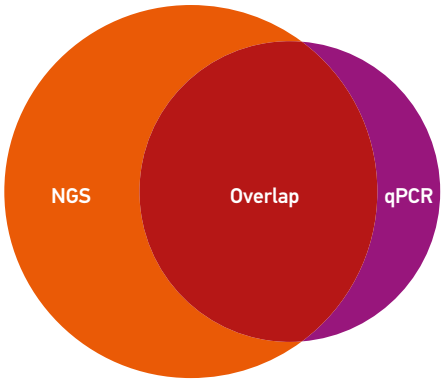
Expression Analysis

Genome-wide comparison of NGS and qPCR platforms for microRNA profiling in serum

Thorarinn Blondal[1], Maurizia Rossana Brunetto[2], Daniela Cavallone[2], Martin Mikkelsen[1], Michael Thorsen[1], Yuan Mang[1], Hazel Pinheiro[1], Ferruccio Bonino[2], Peter Mouritzen[1].
1: Exiqon A/S, Denmark. 2: Digestive and Liver Disease, General Medicine II Unit, University Hospital of Pisa, Italy.

The aim of this study is to compare Next Generation Sequencing (NGS) technologies that have been optimized specifically for biofluid samples, with more established qPCR-based methods for profiling microRNAs in biofluids.

Serum samples from the same patients were analyzed by NGS and qPCR. While there was overall good agreement between NGS and qPCR, there were some differences between the platforms, highlighting the importance of validation.



At a glance

- NGS and qPCR compared: genome-wide microRNA profiling in biofluids
- How to overcome the challenges and achieve reliable results
- How many microRNAs are detected in biofluids
- Overlap between NGS and qPCR results
- IsomiRs and novel microRNAs
- Importance of validating results

Background

Hepatitis B (HBV) and Hepatitis C (HCV) infections are major causes of chronic hepatitis worldwide and chronic infections may eventually develop into cirrhosis and liver cancer (Lavanchy *et al.*, 2008). Improved non-invasive biomarkers are needed to manage patients with chronic HBV and HCV infections, and serum microRNAs may represent a new class of biomarkers due to the important role that microRNAs play in the interaction between virus and host (Fan & Tang, 2014).

In a previous qPCR microRNA profiling study, liver-derived microRNAs were found to be detected at high levels in the sera of HBV patients, and a microRNA signature was discovered that could help identify patients with both natural and therapy induced immune control of chronic HBV infection (Brunetto *et al.*, 2014).

In this study, differences in the serum microRNA profile between HBV and HCV infected patients were investigated.

Overcoming the challenges of Biofluid microRNA profiling

Profiling microRNAs in biofluids is challenging due to the limited amount of RNA present in biofluids, as well as presence of inhibitory compounds which have the potential to inhibit downstream enzymatic processes. In addition, the presence of cellular components may lead to contamination of the "cell-free" biofluid microRNA profile e.g. through hemolysis (lysis of red blood cells).

It is important to standardize sample collection protocols and to monitor any potential sources of pre-analytical variability through rigorous QC procedures.

Each step of the workflow from RNA isolation to RNA QC and NGS library preparation or RT-qPCR needs to be optimized for challenging biofluid samples with limited RNA content.

We have optimized protocols to maximize signals from microRNAs while minimizing carryover of any compounds from biofluid samples, which may inhibit downstream enzymatic processes. Rigorous quality checks have been implemented in every step of the protocols to monitor performance and ensure high quality data.

Methods

Baseline serum samples were collected from 5 patients with chronic HBV infection and 5 patients with chronic HCV infection following standard hospital protocols. Genome-wide microRNA profiling was performed at Exiqon Services on serum from the same individuals by both NGS (Illumina NextSeq500) and qPCR (miRCURY LNA™ Universal RT microRNA PCR System, Human Panels I+II V3) as described in Figure 1.

Different RNA isolation methods have been developed specifically for biofluid NGS and qPCR analysis. Standard methods for measurement of RNA yield and quality, e.g. OD260 measurement, are not reliable for biofluid RNA. Therefore we have developed a qPCR-based method for assessment of RNA quality (Blondal *et al.*, 2013) for which RNA spike-ins are added during the RNA isolation and cDNA synthesis steps.

Different data analysis pipelines have been developed for the NGS and qPCR platforms (Figure 2) to QC and process the raw data prior to performing microRNA differential expression analysis. In the case of NGS analysis, an additional > 50 RNA spike-ins are added during the RNA isolation step, which are later used to monitor the reproducibility and linearity of the sequencing reactions (Figure 3).

Optimized NGS and qPCR methods for biofluids microRNA profiling

NGS	qPCR
RNA isolation from 500 µl serum • Proprietary protocol optimized for NGS microRNA biofluids • Ultra-low elution volumes • Include > 50 RNA spike-ins for RNA QC and sequencing QC	**RNA isolation from 250 µl serum** • miRCURY™ RNA Isolation Kit - Biofluids • Include RNA spike-ins for RNA QC (miRCURY™ RNA Spike-In Kit)
RNA QC • qPCR-based QC to monitor RNA isolation efficiency, inhibition and detect outliers • Hemolysis indicator • Spike-in controls • Endogenous microRNA controls	**RNA QC** • qPCR-based QC to monitor RNA isolation efficiency, inhibition and detect outliers • Hemolysis indicator • Spike-in controls • Endogenous microRNA controls
Library preparation • Proprietary protocol optimized for biofluids with low concentration of starting material (based on NEBNext® Small RNA Library Prep Kit) • Size selection to maximize microRNA reads • QC of library by Bioanalyzer and qPCR	**RT-qPCR** • miRCURY LNA™ Universal RT microRNA PCR System • Human miRNome Panels I + II (742 microRNAs were analyzed) • Protocol optimized for biofluid samples
microRNA Sequencing • Illumina platform (NextSeq500) • 1 x 50 bp reads, 10 M raw reads per sample	

Figure 1. The optimized protocols for NGS microRNA Biofluids are available in Exiqon Services. The optimized protocols for qPCR analysis of microRNAs in biofluids are available in the instruction manuals for the relevant Exiqon products (see Additional Resources). The serum samples were first centrifuged to remove cells and debris, and RNA isolation was performed using 450 µl supernatant (for NGS) or 200 µl supernatant (for qPCR). Further details of the qPCR-based quality control of RNA samples can be found in the Exiqon Biofluids Guidelines.

Analysis pipelines for NGS and qPCR

NGS	qPCR
Raw sequencing data (FASTQ files)	**Raw qPCR data** (Cq values)
Data QC and filtering • Base and read quality • Data quality assessment • Adapter trimming • Identify, remove and analyze > 50 RNA spike-ins (sequencing linearity and reproducibility)	**Data QC** • T_m and melting curve analysis • PCR efficiency • Comparison with negative control (blank purification) • RNA spike-ins and interplate calibrators
Mapping and counting • miRBase 20 • Reference genome • Other reference sources (if applicable) e.g. Rfam, SmallRNA • Abundant sequences (outmapped)	**Filtering** • Flag and remove amplifications that fail to meet acceptance criteria ◦ dissociation curve with single clean peak ◦ amplicon melting temperature consistent for the same assay between samples ◦ PCR efficiency shows no sign of inhibition ◦ sample Cq at least 5 Cq below negative control • Remove assays flagged in > x% of samples* • Remove samples flagged in > x% of assays*
Assembly of novel data • Prediction of novel microRNAs (miRPara) • Mapping to other species in miRBase • Identification of isomiRs	
Apply threshold • TPM threshold • Detected in a group or x% of samples*	**Apply threshold** • Cq value threshold • Detected in a group or x% of samples*
Normalization • Trimmed Mean of M values (EdgeR package)	**Normalization** • Global mean (mean expression value of all microRNAs expressed in all samples after filtering, n = 127 assays)
Differential Expression Analysis • Negative binomial Exact test (EdgeR package)	**Differential Expression Analysis** • ANOVA & Wilcoxon tests

Figure 2. Data quality control and processing applied for NGS and qPCR data from serum samples. NGS data quality assessment includes reports on GC content, Kmer content (stretches of identical bases or over-representation of certain sequence motifs), per base N content (undetermined bases), and per base sequence content. The > 50 RNA spike-ins used for sequencing data QC represent typically 0.1 – 1% of the total number of reads per sample. The thresholds applied to the data are selected depending on the particular project. * Custom analysis offered in Exiqon Services. TPM = Tags Per Million mapped reads.

Results

Platform reproducibility

Both NGS and qPCR platforms displayed excellent technical reproducibility when comparing the results of duplicate sequencing runs on different days from the same library preparation, or duplicate RT-qPCR experiments (Figure 3 and Figure 4).

This confirms that the profiling platforms themselves do not introduce any significant source of variability. Controls such as the > 50 RNA spike-ins (for NGS) and the miRCURY™ RNA Spike-ins and interplate calibrators (for qPCR) are important to monitor the reproducibility between samples in a profiling project.

Excellent technical reproducibility of microRNA sequencing from biofluids

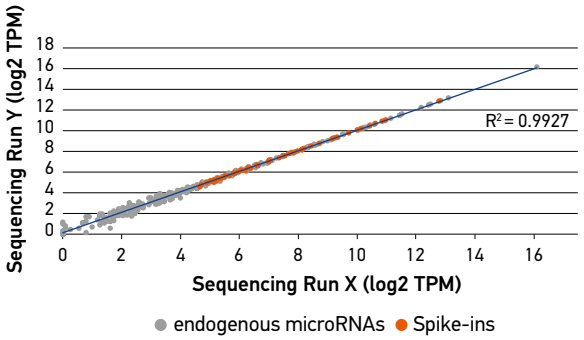


Figure 3. Reproducibility of serum microRNA counts between sequencing runs. The same libraries were sequenced twice on different days using a NextSeq500 instrument. Correlation between two sequencing runs from the same serum sample is shown. Correlation performed using filtered data: only endogenous microRNAs detected > 1 TPM. Endogenous microRNA TPMs are Tags Per Million mapped reads and Spike-in TPMs are Tags Per Million reads.

Excellent technical reproducibility of microRNA qPCR from biofluids

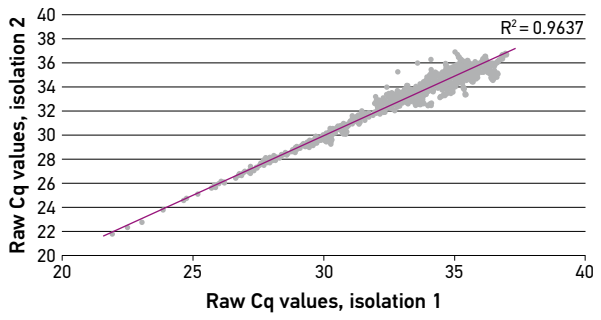


Figure 4. Reproducibility of microRNA Cq values using the miRCURY LNA™ Universal RT microRNA PCR System. The same plasma sample was analyzed twice on different days. Correlation between raw Cq values from different RNA isolations is shown.

A high percentage of mappable reads from serum

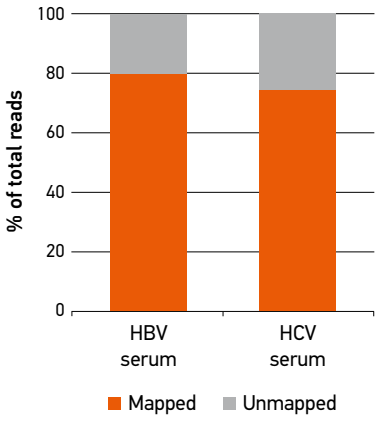


Figure 5. Overall, 80% of the reads obtained for the HBV group of serum samples could be mapped to miRBase or to the human reference genome, assembly GRCh37 (74% for the HCV group). Unmapped reads include sequences not previously reported in the reference genome.

Analysis of microRNA Biofluids NGS reads

A high percentage of the reads obtained from the NGS analysis (74 – 80% on average per group) could be mapped to miRBase or to the human genome (Figure 5). This suggests that the samples and libraries prepared were of good quality, also indicated by the high quality of the sequencing data obtained (base and read Q-scores were > 30 which is equivalent to > 99.9% accuracy).

The NGS libraries were size-selected in order to maximize the reads from the relevant microRNA fraction. However, there was some sample-to-sample variation observed in the percentage of microRNA reads. The HBV serum samples contained on average a higher percentage of microRNA reads than the HCV serum samples (Figure 6).

This difference is attributed to the large amount of liver-derived microRNA present in the serum of HBV patients, where liver derived microRNA were shown to be exported and circulated in the blood within hepatitis B surface antigen (HBsAg) particles (Novellino *et al.* 2012).

How many microRNAs are detected in serum?

The number of microRNAs detected did vary between the different serum samples, and was also influenced by the threshold applied to the dataset (Figure 7). Up to around 900 different microRNAs may be detected by NGS in human serum when no threshold is applied to the data. However, many of these microRNAs are detected at very low read numbers (e.g. below 10 absolute counts).

Low signals in NGS data may represent accurate signals from lowly expressed microRNAs, however small RNA library ligation bias can lead to over-representation of rare sequences, as well as under-representation of abundant sequences (Hafner *et al.*, 2011). Low signals are sometimes difficult to validate due to inherent variability.

The qPCR panels contained assays to detect a maximum of 742 microRNAs. The qPCR data was also subjected to rigorous data QC and filtering (Figure 2), followed by application of a threshold.

Applying a threshold based on Tags per Million mapped reads (TPM) in the case of NGS data, or Cq value in the case of qPCR data, can focus the analysis on the most reliable microRNA signals.

Overall, the threshold applied has the largest impact on the number of microRNAs detected per sample. The total number of microRNAs detected per sample was consistent with results from other serum / plasma projects (comprising patients with non-liver diseases) conducted by Exiqon Services.

Percentage of microRNA reads in serum varies according to biology

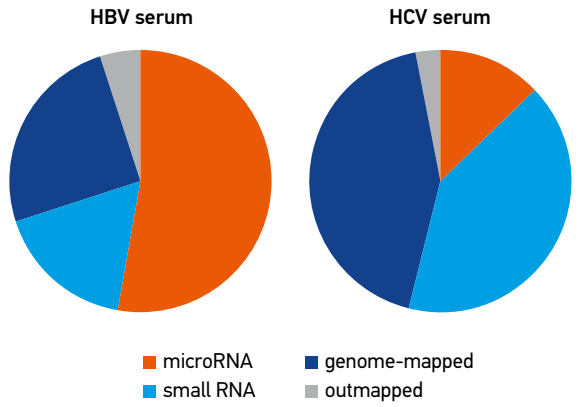


Figure 6. microRNAs in HBV serum samples represented on average 53% of mappable reads, whereas in the case of HCV serum samples the percentage was lower (13%). This is attributed to the large amount of liver-derived microRNA present in the serum of HBV patients. The remaining mappable reads can be categorized as small RNA (of which fragments of tRNAs, Y RNAs, RNUs and Us are the most prevalent), other genome-mapped (degradation products of longer RNAs including mRNA or lncRNA), and outmapped (abundant sequences like mitochondrial and ribosomal RNA as well as homopolymers).

Number of microRNAs detected in serum depends on threshold applied

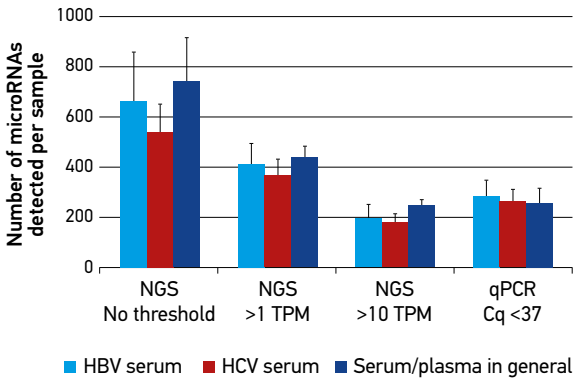


Figure 7. Average number of microRNAs detected per sample by NGS or qPCR. Results are shown for the HBV (n = 5) and HCV (n = 5) serum samples, as well as human serum / plasma samples (from patients with non-liver diseases) profiled in a variety of different projects by Exiqon Services using NGS (n = 165 samples) or qPCR (n = 1,264 samples). Thresholds were applied to the data: values > 1 or > 10 TPM in the case of NGS data, or values < Cq 37 passing filtering in the case of qPCR data. TPM = Tags Per Million mapped reads. Error bars represent the standard deviation.

Which microRNAs are detected by NGS and qPCR?

When comparing the full list of microRNAs detected in any of the 10 samples by NGS or qPCR, there is significant overlap: 348 microRNAs were detected by both platforms (Figure 8).

109 microRNAs were detected only by qPCR. Since the miRCURY LNA™ Universal RT microRNA PCR System has been demonstrated to be the only platform with absolute specificity (Mestdagh *et al.*, 2014), and all qPCR assays have been rigorously wet-lab validated, we are confident that these are valid signals, and not false positives arising from cross-reactivity.

The absence of these 109 microRNAs from the NGS data may be due to small RNA library ligation bias resulting in under-representation of some sequences (Hafner *et al.*, 2011). Alternatively, it may reflect higher sensitivity of qPCR compared to NGS. While platform sensitivity has not been directly compared in this study, this is in agreement with the miRQC study (Mestdagh *et al.*, 2014).

NGS is a hypothesis-free approach, with the possibility to detect any microRNA sequence (within the sensitivity range of the platform), whereas qPCR can detect only those 742 microRNAs included on the qPCR panels. Consistent with this, a substantial number of microRNAs (347) were detected only by NGS.

Overlap in microRNAs detected in serum by NGS and qPCR

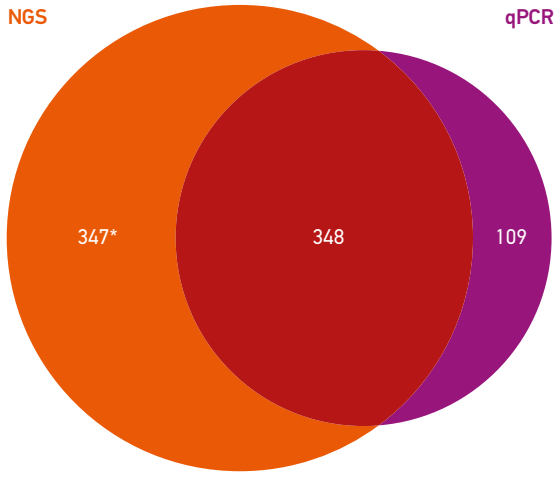


Figure 8. There is significant overlap in the microRNAs detected in serum by NGS and qPCR. microRNAs detected with > 1 TPM or Cq value < 37 in at least 1 of the 10 HBV or HCV serum samples were included. 348 microRNAs were detected both by NGS and qPCR. 109 microRNAs were detected only by qPCR.
* 347 microRNAs were detected only by NGS. The majority of these "NGS only" microRNA are not present on the qPCR panels. qPCR assays are present to detect 60 of the "NGS only" microRNAs, however the qPCR assays failed to detect the microRNAs because they were present predominantly as isomiRs rather than the miRBase sequence.

Comparison of serum microRNA fold changes identified by NGS and qPCR

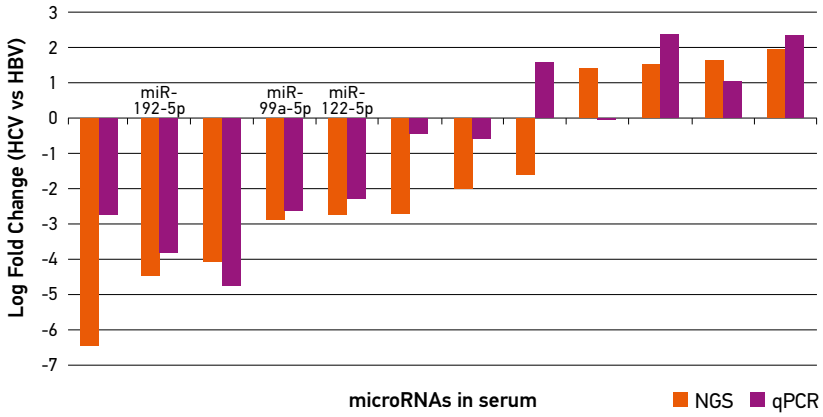


Figure 9. Log fold change in microRNA expression levels between HCV and HBV serum samples show a broad agreement between NGS and qPCR. microRNAs previously identified as differentially expressed in the sera of chronic HBV patients compared with inactive carriers (Brunetto *et al.*, 2014) are indicated by name. A subset of the microRNAs differentially expressed between HCV and HBV are shown. The microRNAs selected are expressed in all samples (> 1 TPM or Cq < 37) and found to be significantly differentially expressed by NGS (p value < 0.05).

IsomiRs and novel microRNAs are detected by NGS

IsomiRs differ in their sequence or length compared to the sequence listed in miRBase. IsomiRs are detected by NGS, but they are poor targets for qPCR assays which are designed to detect the miRBase sequence.

We noted that qPCR assays are present for 60 of the "NGS only" microRNAs, yet these microRNAs were not detected by qPCR. Inspection of the NGS data revealed that the majority of NGS reads resulted from isomiRs of the particular microRNA. Indeed, in several cases, the miRBase sequence was entirely absent from the NGS data.

The NGS analysis (miRPara prediction algorithm) was also able to identify 11 predicted novel microRNAs present in the serum samples, which would require validation using custom miRCURY LNA™ microRNA qPCR assays.

Differential expression consistent between platforms and with published results

The purpose of an expression analysis study is to identify microRNAs which are differentially regulated between sample groups. There was overall good agreement between the NGS and qPCR platforms, with microRNAs differentially expressed between HCV and HBV patients showing similar fold changes (Figure 9). However, a few microRNAs did show opposite trends on each platform, highlighting the importance of careful validation.

Interestingly, certain liver-derived microRNAs (miR-122-5p, miR-192-5p, miR-99a-5p), were found to be detected at lower levels in serum from HCV patients compared to HBV patients. The high level of liver-derived microRNAs in HBV patients is consistent with published results, where the same microRNAs were found at higher levels in the sera of chronic HBV patients compared with inactive carriers (Brunetto *et al.*, 2014).

It is possible that the liver-derived microRNAs may be exported into the bloodstream by defective hepatitis B surface antigen (HBsAg) particles, opening the possibility that around 300 million HBsAg carriers worldwide could be ideal subjects to study the dynamics of liver epigenetics *in vivo*.

Excellent concordance between the miRCURY LNA™ Universal RT microRNA PCR System and NGS platforms was also demonstrated by the miRQC study, which found that 85 – 90% of microRNAs differentially expressed by NGS could be correctly validated with Exiqon's qPCR system (Mestdagh *et al.*, 2014, Figure 10).

Exiqon's qPCR System demonstrates excellent concordance with NGS results

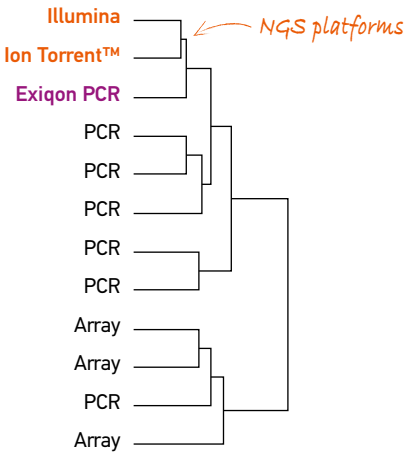


Figure 10. Exiqon's miRCURY LNA™ Universal RT microRNA PCR System is the only one to cluster with the NGS platforms. Concordance of microRNA differential expression status between different NGS, PCR and Microarray platforms. Adapted from Mestdagh *et al.* Nature Methods 2014. Read about the independent miRQC Study: exiqon.com/mirqc.

Conclusions

There is overall good agreement between NGS and qPCR technologies for biofluid microRNA profiling, provided that protocols optimized for biofluids are used.

Results should be validated irrespective of whether the profiling is performed by NGS or qPCR, as there are some differences between platforms.

The choice of platform to use for profiling will depend on several considerations including the goals of the project:

- NGS offers a hypothesis-free approach, whereas qPCR panels are designed to detect a pre-selected set of microRNAs.
- NGS is able to detect isomiRs and novel microRNAs, whereas qPCR assays are designed towards a specific microRNA sequence (usually the one listed in miRBase).
- qPCR may offer higher sensitivity compared to NGS, and does require a smaller amount of biofluid sample.
- NGS profiling on a subset of samples, followed by qPCR analysis of those microRNAs/isomiRs found to be expressed in the samples (using custom miRCURY LNA™ Pick-&-Mix PCR Panels) may provide a powerful combined approach for biofluid microRNA profiling.

Best-in-class Products and Services for Biofluid microRNA profiling

	NGS microRNA Biofluids	microRNA qPCR
Sample type (volume)	• Serum / Plasma (500 µl)	• Serum / Plasma (250 µl)
RNA Isolation	• Exiqon Services proprietary protocol	• miRCURY™ RNA Isolation Kit – Biofluids
Biofluids microRNA profiling	• Exiqon Services proprietary library preparation followed by Illumina sequencing	• miRCURY LNA™ Universal RT microRNA PCR System (Biofluids protocol)

Figure 11. microRNA sequencing using the protocols optimized specifically for biofluid samples is available only in Exiqon Services. qPCR profiling using the protocols optimized for biofluids may be performed in your own lab using the specified products and protocols from Exiqon, or alternatively in Exiqon Services. Serum / plasma is first centrifuged to remove cells and debris, and RNA isolation is performed using 450 µl supernatant (for NGS) or 200 µl supernatant (for qPCR). Products and Services for profiling of microRNAs from other biofluids e.g. urine or CSF and biofluid-derived exosomes are available (see Exiqon Biofluids Guidelines for further information).

Available in Exiqon Services.

References

Blondal *et al.*, Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. 2013 Jan;59(1):S1-6. PMID: 23036329.

Brunetto *et al.*, A serum microRNA signature is associated with the immune control of chronic hepatitis B virus infection. PLoS One. 2014 Oct 28;9(10): e110782. PMID: 25350115.

Fan & Tang. Complex interactions between microRNAs and hepatitis B/C viruses. World J Gastroenterol. 2014 Oct 7;20(37): 13477-92. PMID: 25309078.

Hafner *et al.*, RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA. 2011 Sep;17(9):1697-712. PMID: 21775473.

Lavanchy D. Chronic viral hepatitis as a public health issue in the world. Best Pract Res Clin Gastroenterol. 2008;22(6): 991-1008. PMID: 19187863.

Mestdagh *et al.*, Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. Nature Methods. 2014 Aug;11(8): 809-15. PMID: 24973947.

Novellino *et al.*, Circulating hepatitis B surface antigen particles carry hepatocellular microRNAs. PLoS One. 2012;7(3): e31952. PMID: 22470417.

Additional resources and protocols

Exiqon Biofluids Guidelines:
exiqon.com/ls/Documents/Scientific/microRNA-serum-plasma-guidelines.pdf

Biofluids reading room:
exiqon.com/biofluids

miRCURY™ RNA Isolation Kit – Biofluids instruction manual:
exiqon.com/ls/Documents/Scientific/RNA-Isolation-biofluids-manual.pdf

miRCURY LNA™ Universal RT microRNA PCR instruction manual for biofluid samples:
exiqon.com/ls/Documents/Scientific/Universal-RT-microRNA-PCR-manual-serum.pdf

miRCURY LNA™ Universal RT microRNA PCR, RNA Spike-in kit instruction manual:
exiqon.com/ls/Documents/Scientific/PCR-spike-in-manual.pdf

miRCURY™ microRNA QC PCR Panel instruction manual:
exiqon.com/ls/Documents/Scientific/QC-PCR-Panel-Manual.pdf

Exiqon NGS and PCR Services example reports and guidelines:
exiqon.com/services

Contact information

Outside North America
Phone: +45 45 65 09 29 · Fax: +45 45 66 18 88

North America
Phone: +1 781 376 4150 · Fax: +1 781 376 4152

www.exiqon.com

EXIQON
Seek Find Verify

v1.0 - 03/2014