Exiqon
Services

A guide to the whole transcriptome and mRNA Sequencing Service

Guidelines v1.4
September 2017

EXIQON
Seek Find Verify

Table of Contents

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.
- A copy of the SSF should be sent to Exiqon where it will be reviewed and approved by an Exiqon scientist. Upon approval 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 whole transcriptome or mRNA 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.

Whole transcriptome or mRNA sequencing?

- Exiqon offers two types of long RNA sequencing products: whole transcriptome sequencing and mRNA sequencing. If you are interested in small non-coding RNAs, Exiqon also offer specialized Sequencing services for microRNA (15-30nt) and small RNA (30-200nt), please see **www.exiqon.com/small-rna-ngs**.
- Whole transcriptome sequencing enables the characterization of all RNA transcripts for a given organism, including both the coding mRNA and non-coding RNA larger than 170 nucleotides in length, regardless of polyadenylation. snRNAs and snoRNAs larger than 170 nucleotides in length are also included. However, since ribosomal RNA accounts for vast majority of the transcriptome, the library preparation is designed to remove the ribosomal RNA prior to sequencing. This increases the depth of sequencing for the relevant part of the transcriptome. The library preparation also maintains strandedness information of the RNA transcripts. Stranded information identifies from which of the two DNA strands a given RNA transcript was derived. This provides increased confidence in transcript annotation and enables the detection of antisense transcript expression.
- mRNA sequencing targets all polyadenylated (poly-A) transcripts of the transcriptome. The mRNA library preparation enriches poly-A tailed transcripts from total RNA as part of the library preparation protocol. The poly-A tailed transcripts include the mRNAs (the coding part of the genome), which only accounts for 1-4 % of the whole transcriptome. The enrichment of poly-A tailed transcripts results in an increased depth of sequencing. The increased sequencing depth improves the sensitivity towards identifying lowly expressed mRNA transcripts. The library preparation also maintains strandedness information of the RNA transcripts. This provides increased confidence in transcript annotation.
- Note that mitochondrial poly-A tailed transcripts are considered to be high abundance sequences and are bioinformatically filtered out before mapping takes place.

Comparison of whole transcriptome and mRNA sequencing:

Table 1. Comparison of whole transcriptome and mRNA sequencing

	Whole transcriptome sequencing	mRNA sequencing
RNA species of interest	All RNA transcripts >170 nt	Poly-A transcripts >170 nt
Coding or non-coding	Coding and non-coding transcripts	Coding (Poly-A)
Stranded	Yes	Yes
Enrichment/depletion method	Ribosomal RNA depletion (using biotin-strepavidin based systems)	Poly-A RNA selection (using Oligo dT magnetic bead system)
Differential expression analysis at the gene and transcript level	✓	✓
Identification of splice-junctions	✓	✓
Identification of known transcripts	✓	✓
Prediction of novel transcripts	✓	✓
Antisense transcripts	✓	✓

Number of reads, read length and single or paired end reads?

The depth of sequencing is one of the most crucial factors with regards to both differential expression analysis and discovery of novel transcripts. Guidelines from the ENCODE project (**www.encodeproject.org/ENCODE/protocols/dataStandards/RNA_standards_v1_2011_May.pdf**) state that for a Poly-A enriched experiment 30 million reads is the minimum for most tissues but for detection and analysis of low expression mRNAs over 100 million reads may be needed. Exiqon recommends 50 million reads per sample for most applications. However, if required, it is possible to adjust the number of reads or add extra reads at additional cost. It is also possible to re-sequence the same library to add additional reads later if needed (please contact us if you are interested in this option).

The number of sequencing cycles define the read length. The length of the reads is associated with alignment specificity but longer reads can also affect the overall quality since the quality score (Q-score) drops as the reads get longer. Therefore, Exiqon recommends read lengths of 50bp for most applications. Longer read lengths are available upon request.

In general, paired-end (PE) sequencing (where sequencing is performed from both ends of the molecule) is considered superior compared to single-end sequencing. Paired-end sequencing further decreases the rate of alignment ambiguity which makes transcript assembly more robust. Paired-end sequencing is therefore recommended for discovery applications such as detecting and characterizing novel splice isoforms.

There are a number of caveats associated with comparing libraries of different read length, so it is recommended to use the same read length (and single or paired ends) for all samples that you wish to compare directly.

During your consultation with Exiqon, we will discuss the specific aims of your project and make recommendations regarding the optimal number of reads per sample, read length and single or paired end reads.

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 6).

Total RNA

- We recommend isolating total RNA using a method that retains all RNA species, including small RNAs
- 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 mRNA

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

Tissue and cells

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

Note: RNA isolation aimed at NGS profiling should generally not be isolated with RNA carriers or spike-ins as these may use up many of the reads. Please contact us if you wish to use spike-ins in your experiment.

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 2) shows how much total RNA we recommend to send for whole transcriptome or mRNA 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/ul.

Table 2. Amounts of total RNA needed for whole transcriptome or mRNA QC and sequencing.

	Total amounts of total RNA recommended	Minimum amounts of total RNA recommended	Amount of enriched mRNA recommended
Whole transcriptome sequencing	>300 ng	100 ng + 60 ng for QC	
mRNA sequencing	>300 ng	100 ng + 60 ng for QC	100ng

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

Table 3. Amounts of sample needed for purification and subsequent whole transcriptome or mRNA NGS Services.

Sample type	Amount needed	Comment
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 x10^6 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.
Whole blood	-	Please enquire
FFPE sections	Min 6 x 10 µm sections of 1 cm^2	Not mounted on slides

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 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.

Whole transcriptome or mRNA sequencing is best 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. 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.

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 using a Nanodrop system 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 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.

In general, for sequencing:

- RIN 7-10: Little to no degradation; whole transcriptome or mRNA RNA sequencing possible.
- RIN value 5-7: Partial degradation; whole transcriptome and mRNA sequencing possible. However, mRNA sequencing will be poor since mRNAs will start to lose their 3´ poly-A tails.
- RIN value <5: Degraded RNA; only whole transcriptome sequencing is possible but partial sequencing of degraded rRNA material unavoidable.

Recommended	**Avoid**
1 - High quality RNA	1 – Large differences in RIN values between samples
2 – RNA of similar quality (e.g. all degraded)	2 – One biological group with high quality RNA, one with degraded RNA.

Library preparation and QC

For mRNA Sequencing, poly-A RNA selection is performed using an Oligo-dT magnetic bead system.

For whole transcriptome Sequencing, Ribosomal RNA depletion is performed using biotin-streptavidin based bead systems to minimize ribosomal contamination based on target-specific depletion oligonucleotides. Depending on the sample type and aims of the project, an appropriate depletion method will be selected for your project, resulting in depletion of:

- Cytoplasmic ribosomal RNA or
- Cytoplasmic and mitochondrial ribosomal RNA or
- Cytoplasmic, mitochondrial ribosomal RNA and globin RNA

Following poly-A selection or ribosomal RNA depletion, library preparation is performed including fragmentation, reverse transcription, adapter ligation, pre-PCR amplification (includes addition of sample specific indices).

In Exiqon Services, we perform two types of sequencing library quality controls. Firstly after the library preparation and bead based size selection, the size distribution of the library is evaluated using a Bioanalyzer DNA high sensitivity chip. 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 appropriate flowcell 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 whole transcriptome or mRNA 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

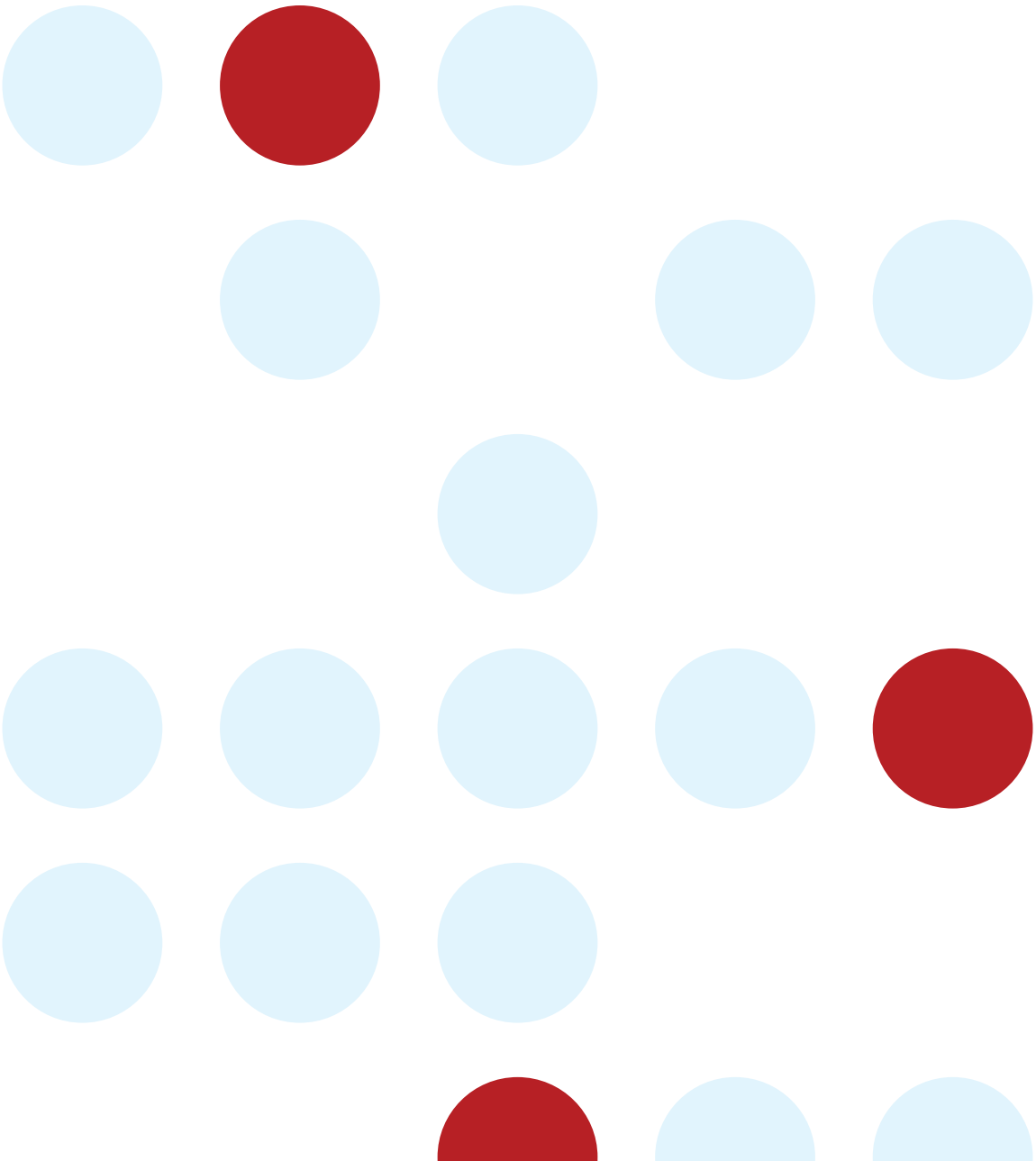
What about customs regulations?

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 includes:

- Comprehensive QC of sequencing data
- Mapping: Alignment of reads to the specified reference genome (see table 4)
- Identification of splice-junctions
- Identification of alternate splicing and splice variants
- Identification of antisense transcripts
- Quantification of known transcripts (both Ensembl and UCSC are supported)
- Prediction and quantification of novel transcripts
- Test for differential expression at gene and transcript level as well as tests for significant changes in promotor usage
- Normalization and group comparison (unsupervised clustering: principle component analysis and heatmap)
- Gene Ontology Enrichment Analysis (GO Analysis)

Specific details are given in the following sections.

Table 4. List of species currently supported (for other species, please inquire).

Species	Common name	Three letter nomenclaure
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

Software tools used for the analysis

Our NGS data analysis pipeline is based on the Tuxedo software package, which is a combination of open-source software, and implements peer-reviewed statistical methods. In addition we employ specialized software developed internally at Exiqon to interpret and improve the readability of the final results.

The components of our NGS data analysis pipeline for RNA-seq include Bowtie2 (v. 2.2.2, see Langmead B and Salzberg S. (2012)), Tophat (v2.0.11, see Trapnell, C., et al. (2009)) and Cufflinks (v2.2.1, see Trapnell, C., *et al.* (2010) and Trapnell, C., *et al.* (2012)), and are described in detail below.

Tophat is a fast splice junction mapper for RNA-seq reads. It aligns the sequencing reads to the reference genome using the sequence aligner Bowtie2. Tophat also uses the sequence alignments to identify splice junctions for both known and novel transcripts as well as identification of insertions and deletions. Cufflinks takes the alignment results from Tophat

and assembles the aligned sequences into transcripts, thereby constructing a map or a snapshot of the transcriptome. To guide the assembly process, an existing transcript annotation is used (RABT assembly). In addition, we perform fragment bias correction which seeks to correct for sequence bias during library preparation (see Kasper *et al*., 2010 and Adam *et al*., 2011).

The Cufflinks assembles aligned reads into different transcript isoforms based on exon usage and also determines the transcriptional start sites (TSSs). When comparing groups, Cuffdiff is used to calculate the FPKM (number of fragments per kilobase of transcript per million mapped fragments) and test for differential expression and regulation among the assembled transcripts across the submitted samples using the Cufflinks output. Cuffdiff can be used to test differential expression at different levels, from CDS and gene specific, down to the isoform and TSS transcript level. For more information on the Cuffdiff module, see Trapnell *et al*., (2013).

QC

As a first step, the read data is subjected to a rigorous QC step to investigate the fidelity of the sequencing data, remove low quality reads and evaluate 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 in order to evaluate mapping percentage and identify reads associated with known transcripts.
- Removal of any remaining mitochondrial sequences, ribosomal RNA sequences, poly-A / poly-C homopolymer sequences and other unwanted species.

Mapping

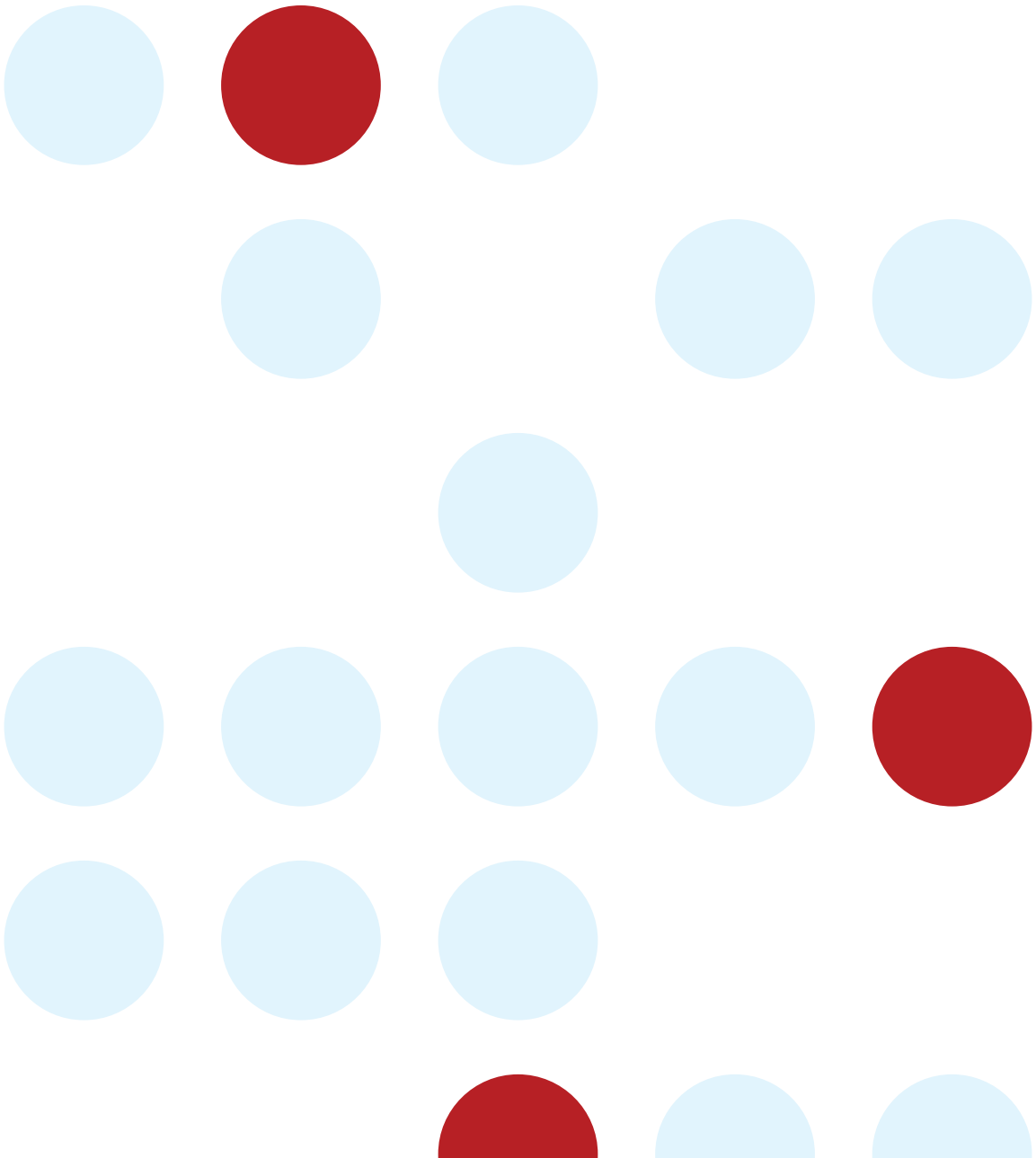
- Alignment to reference genome (species supported at standard price are listed in table 4)
- Splice junctions are identified (where relevant)
- Quantification of known transcripts
- Detection of antisense transcripts
- Identification of novel transcripts
- A table of known and novel transcripts and their associated number of reads is then generated for differential expression analysis. This forms the whole transcriptome or mRNA sequence profile for the sample

Normalization and differential expression analysis

- Normalization attempts to correct for effects caused by variation in sequencing depth between samples. We use the standard Fragments Per Kilobase of exon per Million mapped reads (FKPM) implemented in Cufflinks to perform the normalization of datasets.
- The differential expression analysis investigates the relative change in expression (i.e. reads) between features in different samples. Differential expression analysis is performed both at the gene and at the transcript level based on read counts . The analysis is performed by Cuffdiff and attempts to distinguish true differentiation effects from noise by estimating effects from factors such as biological variation between replicates, count uncertainty (caused by ambiguous mapping of reads) and count overdispersion (caused by greater than expected variation in reads amongst replicates).
- A list of transcripts predicted to be differentially expressed between different experimental groups is presented in the final report.In addition we provide groupwise comparisons (unsupervised clustering: principle component analysis and heatmap).

Biological interpretation of results

To aid interpretation of the results, we perform a final step where we provide a rudimentary biological interpretation of the findings. We perform a Gene Ontology (GO) enrichment analysis to determine which GO terms are overrepresented in the differentially expressed transcripts.



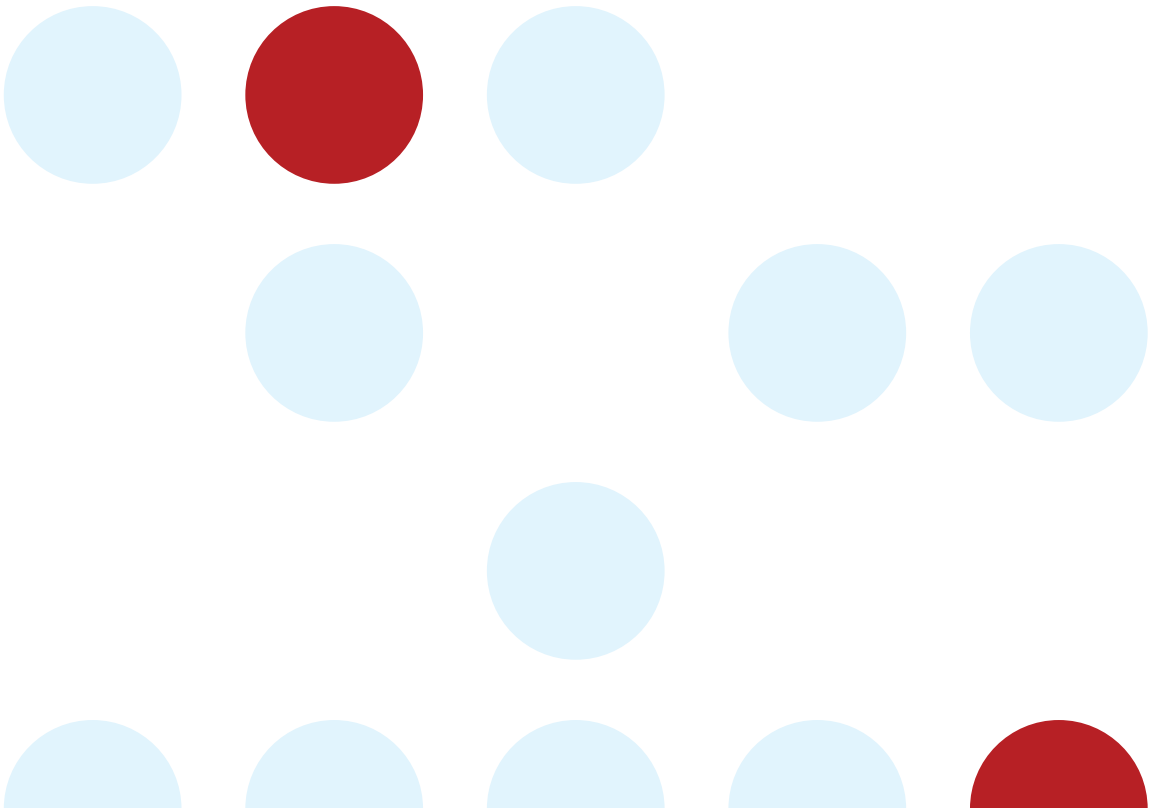
Report and final consultation

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. This includes:

- 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. The report will include process, QC and data reporting with overview, counting, normalization and differential expression analysis.
- Extensive Spreadsheet files with all the major findings and statistical analysis.

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

- We supply the raw FASTQ files, BAM files and data analysis output files (.TSV) for the aligned reads on an encrypted storage hard-drive.
- Exiqon offers a free consultation with one of our RNA NGS experts to discuss the data, to answer questions you may have, and discuss the next steps for the project such as qPCR validation or functional analysis.
- For standard projects please allow approximately 4-6 weeks for completion of the entire analysis from RNA samples passing QC at Exiqon 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.
- 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.
- The unused portion of any samples will be discarded 3 months after completion of the experiments unless specifically agreed otherwise.



Frequently asked questions

Question: Can Exiqon perform sequencing on species other than the species listed in table 4?
Answer: Please contact us to discuss projects involving other species and we are happy to set-up a custom data analysis for you.

Question: Can Exiqon perform sequencing on bacterial or viral samples?
Answer: Yes. please contact us for a detailed discussion. Please note that if viral transcriptome analysis is to be performed both the host and the virus reference genomes need to be available.

Question: Which gives the greatest depth of coverage of messenger RNAs - whole transcriptome or mRNA Sequencing?
Answer: In general, mRNA sequencing enriches the polyadenylated transcripts, hence the reads are focused on a smaller number of transcripts which can therefore be sequenced to a greater depth. Consequently, mRNA sequencing offers the greatest depth of coverage for mRNAs compared to whole transcriptome. However, the depth of coverage also depends on the total number of reads in the experiment, and we can discus this topic further during your initial consultation to find out which service would best meet the aims of your project.

Question: Which service should I choose if I am interested in long non-coding RNAs?
Answer: Long non-coding RNAs will be included in the whole transcriptome sequencing service. Some long non-coding RNAs may be captured in the mRNA sequencing service, but only those which are polyadenylated. Hence, mRNA sequencing is not suited for detection of long non-coding RNAs.

Question: Can the whole transcriptome sequencing service be extended to include small non-coding RNAs less than 170nt in length?
Answer: Small fragments may be poorly represented in the library due to the way the libraries are generated for whole transcriptome sequencing (random priming). Therefore for sequencing of small RNAs we recommend to combine with our microRNA or small RNA sequencing service.

Question: Are circular RNAs included in the whole transcriptome sequencing?
Answer: If the circular RNAs break down to generate fragments over 170 nt in length they will be included in the library.

Question: Are capped RNAs included in the whole transcriptome or mRNA sequencing?
Answer: Yes, since transcriptome and mRNA sequencing is based on random priming to generate cDNA, 5'capped RNAs will be included.

Question: What is the minimum RNA amount needed for whole transcriptome or mRNA Sequencing?
Answer: It depends on the application and sample type. Please refer to table 2, page 6 for recommended input amounts. Please contact us if you have limited amounts of RNA available from your samples, and we will be happy to advise.

Question: Do I need to perform a DNase treatment before sending the total RNA to Exiqon for sequencing?
Answer: Since the library generation is based on addition of double stranded adaptors onto repaired A-overhang double stranded cDNA, DNA contamination can interfere and cause contamination. Make sure to use RNA isolation method that depletes DNA.

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 the data will be affected by degraded RNA, leading to higher quality data. We recommend a RIN value > 7 for whole transcriptome and mRNA sequencing. Note: lower RIN values will result in a greater portion of the data coming from degraded RNA which will result in a decreased sequencing depth of relevant transcripts. Also, uneven coverage of mRNA transcripts (enrichment of the 3' end) can occur if doing mRNA seq on degraded samples.

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 transcriptome 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.

Question: Can you analyze pre-prepared libraries?
Answer: Please contact us if you are interested in sequencing libraries you have already prepared. We will advise on the requirements.

Question: How does Exiqon select the polyA RNA prior to mRNA sequencing?
Answer: Exiqon uses the Illumina TruSeq Stranded mRNA Kit which captures the poly-A transcripts using Oligo-dT.

Question: How many reads will I obtain per sample?
Answer: The number of reads per sample depends on the sequencing instrument being used, and how many samples are multiplexed together. For example, when using the HiSeq 2500 we typically obtain around 250M reads per lane (which is divided between the samples being multiplexed together), and there are 8 lanes in total. Exiqon will determine the number of samples being multiplexed together, based on the number of reads required for your project.

Question: Are these reads "guaranteed" reads?
Answer: The reads for a given project are average raw reads that pass quality filtering (CHASTITY filtering), with a variation of 10-15% per sample. If you order 50 M reads per sample, you can expect your samples to have 42.5-57.5 million reads.

Question: Can I do accurate quantification with 50 million reads?
Answer: 50 million reads are sufficient for most applications. Project specific details can be discussed during an initial consultation

Question: The product offering says 50 million reads. However when looking at my data only 42 million map to transcripts
Answer: We generate 50 million passed filter reads for each sample with a variation of 10-15%. Normally, of these, typically 65-90% of the reads are mapped to the reference genome, depending on the RNA quality, type of sample and type of enrichment/depletion method.

Question: What does strandedness mean?
Answer: With directional RNA sequencing, the original direction of the biological material for every read is preserved. Stranded information identifies from which of the two DNA strands a given RNA transcript was derived. This provides increased confidence in transcript annotation and enables the detection of antisense transcript expression. Using a strand-specific protocol, it is possible to accurately quantify overlapping transcripts that are transcribed on different strands. This would not be possible with a non-directional protocol.

Question: Is the RNA fragmented prior to whole transcriptome or mRNA sequencing?
Answer: Yes, after depletion of ribosomal RNA or polyA selection, the RNA is fragmented enzymatically prior to the library preparation.

Question: What instruments will be used for the whole transcriptome or mRNA sequencing?
Answer: Exiqon Services uses a range of sequencing instruments from Illumina, including the Illumina NextSeq 500 and HiSeq 2500.

Question: Will you be barcoding and multiplexing samples together?
Answer: All NGS libraries will be made individually and can therefore be compared individually as well. However to save cost 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: Will barcodes be introduced during the PCR or in the adaptors?
Answer: The indices are added during the PCR as part of the 3'-primer.

Question: Do Exiqon map the reads to both the known genome sequence as well as known transcripts?
Answer: Yes, we map the reads against the relevant annotated reference genome. This means that we will do an initial mapping of the sequenced reads agains the part of the genome which contains known genes and transcripts (features). The reads which do not align to these parts of the genome are used for identification of novel transcripts.

Question: How are novel transcripts and lncRNAs identified?
Answer: The reads which do not map to known genes or transcripts are used to predict novel transcripts. These transcripts can encompass both novel gene isoforms or lncRNAs.

Question: What databases do Exiqon use to map lncRNA?
Answer: We support reference annotations from both Ensembl and UCSC.

Question: How do Exiqon identify splice junctions and alternative splice forms?
Answer: The process of decomposing aligned reads into a set of transcripts is based upon a peer-reviewed method described in "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation", Nature Biotechnology, 28,511–515 (2010) doi:10.1038/nbt.1621

Question: Does Exiqon's analysis include identification of SNPs?
Answer: We do not currently perform SNP identification.

Question: Which reads are removed by data quality controls?
Answer: Reads need to have a high quality score (Q score) and a (trimmed) read length >35. If the data points do not meet these criteria, they are removed from the dataset.

Question: What are 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 reduce 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. GO enrichment analysis is included in the data analysis from Exiqon.

Question: Is KEGG analysis included?
Answer: This is currently not part of Exiqon's standard data analysis package. Please enquire if you would be interested in this as a customized bioinformatics solution.

Question: Which normalization procedures are performed on whole transcriptome or mRNA NGS data?
Answer: For transcriptome analysis we use Fragments Per Kilobase of exon per Million mapped reads (FPKM) implemented in Cuffdiff.

Question: What data filtering is performed? Are rRNA reads filtered out?
Answer: The raw sequence reads are first filtered to remove adaptor sequences. Subsequently we remove all reads which align to unwanted sequences, e.g rRNA and mitrochondrial RNA. The remaining reads are then aligned to the reference genome.

Question: Can Exiqon provide an integrated analysis of microRNA and messenger RNA sequencing data?
Answer: This is currently not part of Exiqon's standard data analysis package. Please enquire if you would be interested in this as a customized bioinformatics solution.

Question: Can Exiqon provide a list of microRNAs known to regulate the messenger RNAs identified as differentially expressed in my dataset?
Answer: This type of information is available through our freely available miRSearch database (www.exiqon.com/miRSearch).

Question: Is it possible to add additional samples at a later date?
Answer: Yes, it is possible to do this. Additional bioinformatics analyses can be requested as a custom service.

Question: What if I would like to perform the analysis myself? Can Exiqon provide only the raw data?
Answer: Yes, data will be passed through the initial QC pipeline to ensure that the data meets our quality standards. We will then send the raw FASTQ files on an encrypted hard disk.

Question: I am interested in validating several novel and known transcripts. What do I do?
Answer: Exiqon offers validation of NGS results on Pick-&-Mix qPCR panels including custom designed LNA qPCR assays for any mRNA or lncRNA transcript: http://www.exiqon.com/pick-and-mix. However, it is important to consider the read counts of the transcript of interest. Transcripts with low read counts may be hard to validate.

Question: how do we perform fragmentation of the RNA?
Answer: Fragmentation is done enzymatically

Question: If I want FASTQ files only, is there any QC done on the data?
Answer: There will be an independent QC on the raw reads apart from the QC report coming from the instrument. This is a standard QC step which doesn't cost extra.

Question: what is the difference between single read (SR) and paired end read (PE) and the effect on my data?
Answer: The advantage of paired-end reads compared to single-end reads is that paired-end sequencing is more likely to map unambiguously to the reference genome. This improves the ability to assemble transcripts, detect rearrangements, including insertions and deletions and inversions. However, the paired-end sequencing also increases the cost of the sequencing but the library preparation protocol is the same as for single reads sequencing.

Question: Is it possible to do RNASeq specifically for lncRNA? Could we deplete the mRNAs somehow?
Answer: No, we do not offer that service.

Question: Can you explain why this size cut-off (170nt)? Why do I not get the small ncRNA too?
Answer: That is due to the the random priming during the DNA synthesis, the smaller fragments (<170) will be poorly represented.

Question: Can you provide integrated microRNA and mRNA analysis?
Answer: The two sequencing protocols are done with different protocols and are run in separate pipelines and we do not offer this in our standard setting. Please consult us for custom bioinformatics service.

Question: Is there a difference between the data analysis for the Whole transcriptome and mRNA seq?
Answer: No, it's the same pipeline based customized but based on the Tuxedo suite. We make no distinction between reads coming from either sequencing method.

Question: So what does 1 FPKM really mean in terms of abundance that is copies of RNA?
Answer: This is difficult to estimate and highly variable from cell type to cell type and the total number of mRNAs in a given cell. Marinov et al. states for example "We also estimate that for GM12878 single cells, one transcript copy corresponds to 10 FPKM on average. This agrees well with the observation that detection of genes becomes unstable below ;100 FPKM, which is also consistent with previous observations (Ramskold, et al. 2012).
Kellis et al. states that "FPKMs are not directly comparable among different subcellular fractions, as they reflect relative abundances within a fraction rather than average absolute transcript copy numbers per cell. Depending on the total amount of RNA in a cell, one transcript copy per cell corresponds to between 0.5 and 5 FPKM in Poly-A+ whole-cell samples according to current estimates (with the upper end of that range corresponding to small cells with little RNA and vice versa".

Question: What is a "novel" RNA transcript?
Answer: A novel transcript is characterized as a transcript which contains features not present in the reference annotation. Identification of novel transcripts depends therefore upon the reference annotation. Currently we support the annotations provided in Illuminals iGenome Ensembl and UCSC downloads.Thus, transcripts not in these annotations will be classified as novel transcripts.

Question: A novel transcript identified seems to be a known gene when I look it up in the gene browser, why is that?
Answer: Most novel transcripts are not new "genes" but different isoforms of already annotated genes. Most likely it has a novel combination of exons or a different start site.

Terminology

FASTA and FASTQ	FASTQ is a format for storing sequence data along with quality scores from a sequencing run. The quality format uses a single ASCII character per location. It can be opened in standard text editors. FASTA is the same format without the quality scores. Note that large files can be difficult to handle by the editor. *For example, a FASTQ sequence can be:* SEQ_ID_XYZ CATACGCGCTATAGCGCGATAGAGCTCTCGATGTATGGGTATACG + !+))%%%++)(%%%*).1***-+*''**55F>>>c>CCC%CC65 *While the same sequence in a FASTA format is:* SEQ_ID_XYZ CATACGCGCTATAGCGCGATAGAGCTCTCGATGTATGGGTATACG
SAM	Sequence Alignment/Map format. Contains mapping information of each read from a sequencing run along with the sequence information. Is readable in standard text editors. However it is recommend to use software packages such as BEDtools or samtools to access the data.
BAM	Binary Alignment/Map format. A binary version of the SAM file that allows compression of the data to save space on the computer. Can be accessed through BEDtools or similar softwares.
FPKM	Fragments per kilobase cDNA (or transcript) per million mapped reads
CDS	The whole coding sequence/region, known to be translated into a protein
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.

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.

References

Goff L., *et al.* (2012) http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html.

Kasper D., *et al.*(2010), Biases in Illumina transcriptome sequencing caused by random hexamer priming Nucleic Acids Research, Volume 38, Issue 12.

Kellis, M., *et al.* (2013) Defining functional DNA elements in the human genome. PNAS early edition, April 21, 2014.

Langmead, B., *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10(3): R25.10.

Langmead B, Salzberg S. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359.

Marinov, G. K., *et al.* (2014) From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and RNA splicing. Genome Res. 24: 496-510.

Ramskold, D., *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30: 777–782.

Roberts, A., *et al.* (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics, 27(17): 2325-2329.

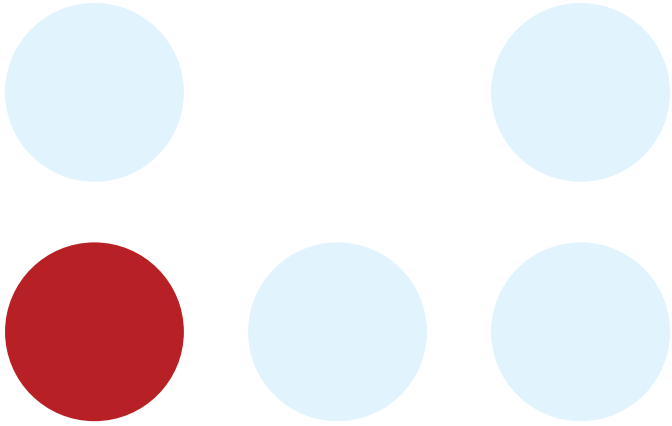
Roberts, A., *et al.*, (2011) Improving RNA-Seq expression estimates by correcting for fragment bias Genome Biology, Volume 12, R22.

Trapnell, C., *et al.* (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England), 25(9):1105-1111.

Trapnell, C., *et al.* (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5): 511-515.

Trapnell,C., *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7,562–578.

Trapnell, C., *et al.* (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology 31, 46-53.



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

923004 - v1.4 - 09/2017