Microsynth

Genome-wide Profiling of microRNAs by Next Generation Sequencing

Find out how gene expression is regulated
Allows inferring functional regulatory effects

Introduction

microRNAs (miRNA) are a class of small non-coding RNAs typically 21-23 nt long found in plants, animals, and some viruses (see ***Figure 1***). miRNAs play a pivotal role in RNA silencing and post-transcriptional regulation of gene expression. Discovered in the early 1990s, miRNA research has revealed (i) multiple roles for miRNAs in development (ii) disease-associated aberrant expression of miRNAs and (iii) the importance of miRNA in many other biological processes.

Next generation sequencing (NGS) technologies have become a powerful tool to study genome-wide miRNA expression patterns and have helped to identify disease associations, isoforms of miRNAs, and to discover previously uncharacterized miRNAs.

Figure 1. *Typical structure of two precursor miRNAs showing the stem loop structure observed in precursor miRNAs (pre-miRNAs). The primary miRNA transcript which varies between 500 and 3000 nt is processed by RNAse III and the dsRNA binding protein, resulting in a 70-80 nt long pre-miRNA. The pre-miRNA is then actively transported from the nucleus to the cytoplasm where it is further processed by the protein Dicer resulting in mature 17-23 nt long miRNAs.*

Microsynth Competences and Services

Experimental Design: As an expert in the area of miRNA-Seq, Microsynth is able to provide a one-stop service from experimental design consulting up to bioinformatics analysis (see ***Figure 2***). In case you do not involve Microsynth in your experimental design, please consider the importance of the number of biological replicates. We usually advise including at least 3 biological replicates per condition in order to finally obtain statistical significance for your differential miRNA expression analysis.

RNA Isolation: Either you leave it up to Microsynth or you use a commercial kit to isolate total RNA used for the Illumina miRNA-Seq protocol.

Library Preparation and Sequencing: Following a quality check of your total RNA samples, Microsynth will perform miRNA enrichment. Illumina cDNA library is generated by reverse-transcription including specific sequencing adaptors with barcodes. Finally, the libraries are pooled and sequenced on the Illumina machine. The envisaged number of reads per library depends on the organism under study and the desired sensitivity. The usually required number of reads for higher eukaryotic species (e.g. human, rat, mouse) is approx. 5-15 million reads, depending on whether complex tissues or unique type of cells are analyzed.

Bioinformatics Analysis: The analysis pipeline at Microsynth addresses three main questions: (i) what is the distribution of miRNAs and which of them are novel, (ii) which pathway is influenced in which way by the miRNAs and (iii) which of the miRNAs are differentially expressed. The first step of analysis is based on the sequence data itself. In short, sequence data is quality filtered and clustered for each condition of the

Microsynth AG, Switzerland
Schützenstrasse 15 · P.O. Box · CH-9436 Balgach · Phone +41 71 722 83 33 · Fax +41 71 722 87 58 · info@microsynth.ch · www.microsynth.com

experiment. A representative sequence of each cluster is then compared against the miRBase database using UBLAST to identify known miRNAs. Sequence clusters that did not result in a significant hit may be regarded as putative novel miRNAs (see ***Figure 3***). In the second step of the analysis the quality filtered reads are mapped against the reference genome using STAR. Then, HOMER is used to find miRNA peaks and motifs and to exhaustively annotate them (see ***Figure 4***). However, this in-depth annotation is only supported for a limited set of model organisms (e.g. human, mouse, zebra-fish). Finally, differentially expressed miRNAs are found using DESeq2 (see ***Figures 5 - 6***).

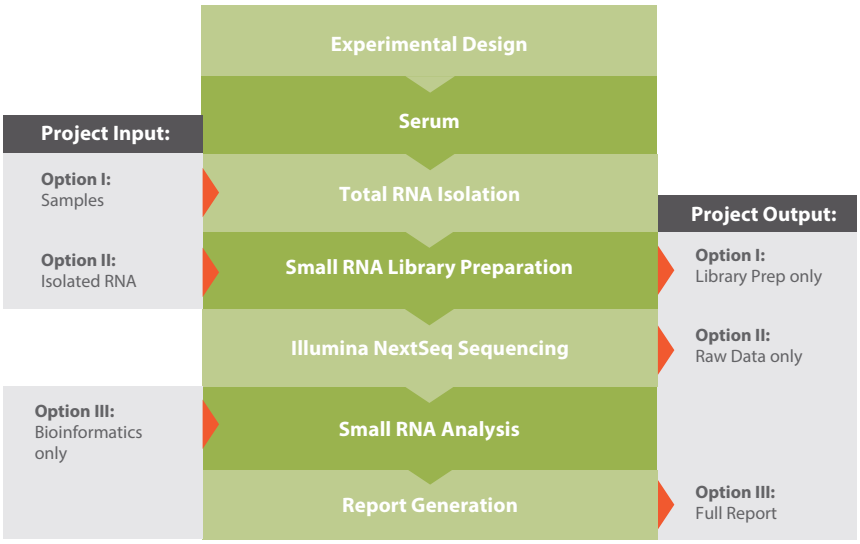

Figure 2. *Microsynths workflow for small RNA projects. The workflow can be entered and exited at different points depending on customer requirements.*

Examples for the Most Important Output Files Provided by Microsynth

query id	subject id	% identity	alignment length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
Sample1_R1.3;size=1311696;	mmu-miR-122-5p	100.0	22	0	0	1	22	1	22	2.7e-07	41.7
Sample1_R1.2;size=1456590;	mmu-miR-122-5p	100.0	21	0	0	1	21	1	21	9.2e-07	39.9
Sample1_R1.1;size=1824665;	mmu-miR-21a-5p	100.0	21	0	0	1	21	1	21	9.2e-07	39.9
Sample1_R1.8;size=518165;	mmu-let-7f-5p	100.0	21	0	0	1	21	1	21	9.2e-07	39.9
Sample1_R1.5;size=793584;	mmu-miR-122-5p	100.0	20	0	0	1	20	1	20	3.2e-06	38.1
Sample1_R1.7;size=625497;	mmu-miR-21a-5p	100.0	20	0	0	1	20	1	20	3.2e-06	38.1

Figure 3. *Exemplary extract of the comparison of cluster representative sequence (query id) against the miRBase database (subject id) using UBLAST. Clusters that show no significant blast hit can be regarded as putative novel miRNAs.*

Rank	Motif	p-value	log p-value	q-value/FDR	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-22	-5.173e+01	0.010	20.37%	0.07%	27.9bp (53.3bp)	NA More Information	motif file (matrix)
2		1e-10	-2.370e+01	0.060	7.41%	0.01%	43.4bp (18.7bp)	NA More Information	motif file (matrix)
3		1e-9	-2.264e+01	0.070	9.26%	0.05%	35.1bp (26.9bp)	NA More Information	motif file (matrix)
4		1e-9	-2.126e+01	0.055	55.56%	17.88%	40.0bp (54.6bp)	NA More Information	motif file (matrix)
5 *		1e-7	-1.762e+01	0.300	11.11%	0.31%	38.8bp (52.4bp)	NA More Information	motif file (matrix)

Figure 4. *Exemplary extract of the de-novo motif search results. Given the p-value and the false discovery rate, motif 5 (*) is rated as putative false positive motif. Additionally total target and background sequences are listed and links to known motifs are provided if present (not shown).*

10 records per page

Search all columns:

ID	Image	From to logFC	From to p-Value	From to Adjusted p-Value
mmu-miR-148a-3p		-3.000000	5.47e-41	2.84e-38
mmu-miR-148b-3p		-2.530000	4.87e-34	1.27e-31
mmu-miR-101a-3p		-2.620000	1.82e-33	3.15e-31
mmu-miR-148a-5p		-2.460000	2.28e-28	2.96e-26
mmu-miR-143-3p		-1.970000	3.55e-22	3.69e-20
mmu-miR-92a-3p		-2.070000	3.21e-18	2.62e-16
mmu-miR-28a-3p		-1.960000	3.52e-18	2.62e-16
mmu-miR-345-3p		-2.110000	6.49e-16	4.22e-14
mmu-miR-125b-2-3p		-1.870000	2.95e-14	1.70e-12
mmu-miR-532-5p		-1.640000	7.23e-13	3.76e-11

Figure 5. *Exemplary extract of differentially expressed miRNAs depicting their distribution (boxplots) and listing their corresponding statistics.*

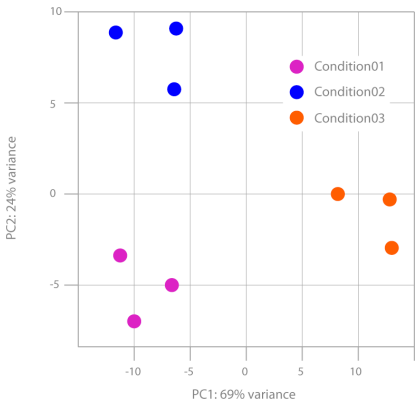

Figure 6. *Differential miRNA expression for three different experimental conditions, whereby each condition includes three replicates. The expression data was submitted to principal component analysis (PCA) to show differences among replicates and conditions.*

Related Topics

- siRNA synthesis service at Microsynth
- RNA-Sequencing at Microsynth
- ChIP Seq analysis pipeline at Microsynth

Further Reading

1. *Griffiths, J.S., (2004) The microRNA Registry. Nucl. Acids Res. 32 (suppl 1): D109-D111.*
2. *Dobin et al. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15-21.*
3. *Love et al (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550.*
4. *Edgar R.C. (2010) Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26: 2460-2461.*
5. *Heinz et al. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 38: 576-589.*
6. *Mestdagh et al. (2014) Evaluation of quantitative miRNA expression platforms in the microRNRNA quality control (miRQC) study. Nat Methods. 11: 809-815.*