Microsynth

De novo Transcriptome Assembly and Annotation

Study the transcriptome of any organism

Discover differential gene expression whithout a reference genome

Introduction

RNA sequencing by next generation sequencing (NGS) has become a standard tool for studying the transcriptome of an organism and related gene expression profiles. A mandatory prerequisite however, is the availability of a well-annotated reference genome or transcriptome. This reference is used to map the sequencing reads, to count the reads mapped to genetic features (e.g., genes, transcripts) and to perform statistical analysis by comparing sample replicate groups from different conditions to each other. If such a reference is missing, no differential gene expression analysis will be possible unless a reference transcriptome is built. Microsynth's *de novo* transcriptome assembly and annotation service is designed to fill this gap and allow the study of yet unknown transcriptomes.

Microsynth's Competences and Services

With more than ten years of experience in the field of next generation sequencing, one of Microsynth's core competences is to provide high-quality one-stop services from experimental design to bioinformatics analysis. You may either outsource the entire analysis or only single steps to us as illustrated in ***Figure 1***.

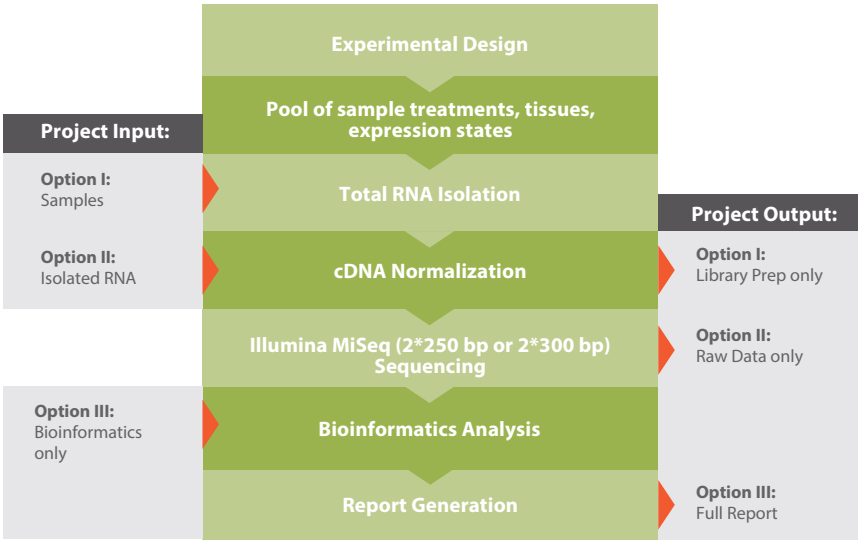

Figure 1. *Microsynth's workflow for de novo transcriptome projects. The workflow can be entered and exited at various steps depending on the requirements of the customer.*

Experimental Design

Microsynth's NGS specialists will help you define suitable experimental setups for your *de novo* trancriptome project and discuss possible strategies to address your associated research questions.

RNA Isolation

You may either perform the RNA extraction yourself or outsource this critical step to Microsynth, which has long-standing experience in processing various sample matrices and RNA sources.

Library Preparation and Sequencing

To be able to cover the transcriptome of an organism well enough, RNA from different tissues, stages of growth and reactions to treatments need to be pooled and sequenced together. This recommendation is in stark contrast to the standard differential expression analysis in which sample groups must not be pooled together to be able to detect variations in gene expression patterns. From the extracted RNA, a normalized cDNA library is created to represent every transcript as uniformly as possible for the future reference transcriptome. Sequencing is performed

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

on one of our Illumina MiSeq platforms which may produce paired-end reads with up to 2x300 base pairs.

Bioinformatics Analysis

Sequenced reads are quality filtered, trimmed and *in silico* normalized before being subjected to *de novo* transcriptome assembly. The resulting assembly is further refined, and transcripts are predicted from the assembled contigs. The transcripts are then searched against the Swissprot and PfamA databases for homologies and are annotated with any available supplementary information such as GO and KEGG terms. A detailed description of the data- and workflow used by Microsynth to create reference transcriptomes can be found in the following publication [1]. ***Tables 1*** and ***2*** are visual cues on how the results are presented.

Example Results

Table1. *The detected transcripts are described in detail in a GFF3 formatted file along with their untranslated (UTR) and coding (CDS) regions as depicted in this cutout.*

TRINITY_DN16040_c1_g2_i13	transdecoder	gene	1	1484	.	-	.	ID=TRINITY_DN16040_c1_g2~~TRINITY_DN16040_c1
TRINITY_DN16040_c1_g2_i13	transdecoder	mRNA	1	1484	.	-	.	ID=TRINITY_DN16040_c1_g2_i13.p1;Parent=TRINITY
TRINITY_DN16040_c1_g2_i13	transdecoder	five_prime_UTR	660	1484	.	-	.	ID=TRINITY_DN16040_c1_g2_i13.p1.utr5p1;Parent=
TRINITY_DN16040_c1_g2_i13	transdecoder	exon	1	1484	.	-	.	ID=TRINITY_DN16040_c1_g2_i13.p1.exon1;Parent=T
TRINITY_DN16040_c1_g2_i13	transdecoder	CDS	339	659	.	-	0	ID=cds.TRINITY_DN16040_c1_g2_i13.p1;Parent=TRI
TRINITY_DN16040_c1_g2_i13	transdecoder	three_prime_UTR	1	338	.	-	.	ID=TRINITY_DN16040_c1_g2_i13.p1.utr3p1;Parent=

Table 2. *A cutout of one of the produced annotations detailing homologies of the assembled transcripts to sequences found in various publicly available databases:*

Transcript	UniProtKB-AC	UPKB-ID	uptarget descr	% identity	alignment length	mismatches	gap opens	evalue	bit score	PfamAC	pftarget name	pftarget descr	target len	query len	evalue	score
TRINITY_DN16040_c1_g2_i13.p1	O22977	BDG3_ARATH	Probable lysophospholipase BODYGUARD 3 OS=Arabidopsis thaliana OX=3702 GN=BDG3 PE=2 SV=1	62.8	86	28	1	6.4e-27	121.3							
TRINITY_DN16040_c1_g2_i13.p1										PF00561.19	Abhydrolase_1	alpha/beta hydrolase fold	256	107	0.0028	17.3
TRINITY_DN22587_c1_g1_i3.p1	Q9ZPV7	RER1D_ARATH	Protein RER1D OS=Arabidopsis thaliana OX=3702 GN=At2g18240 PE=2 SV=2	72.3	112	31	0	5.2e-43	175.3							
TRINITY_DN22587_c1_g1_i3.p1										PF03248.12	Rer1	Rer1 family	171	148	2.1e-47	161.0

Related Topics

1) For Microsynth's standard RNA sequencing, differential gene expression and pathway analysis services, please visit the related website and download the application note.
2) For Microsynth's gDNA *de novo* assembly service, please visit the following website and download the application note.

References

[1] Ricardo C. Neves, Joao C. Guimaraes, Sebastian Strempel, Heinrich Reichert. Transcriptome profiling of Symbion Pandora (phylum Cycliophora): insights from a differential gene expression analysis. Org Divers Evol. 2017 17:111-119. DOI 10.1007/s13127-016-0315-1. http://rdcu.be/m8lt