Upload image

Version 4

Aug 05, 2020

Using sequins with RNA sequencing. V.4

Tim Mercer[1]

[1]Garvan Institute of Medical Research

1 *Works for me* This protocol is published without a DOI.

Tim Mercer
Garvan Institute of Medical Research

ABSTRACT

Sequins are synthetic RNA controls that that are 'spiked-in' to your RNA sample, and undergo concurrent library preparation, sequencing and analysis. The sequins are then analyzed as internal controls in the output NGS library.

This protocol describes the laboratory steps required to dilute, store and spike the sequins into your RNA sample prior to library preparation for RNA sequencing. We also describe the bioinformatic steps to analyse sequins within your read (.FASTQ) or alignment (.BAM) files.

EXTERNAL LINK

www.sequinstandards.com

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Hardwick et. al., Spliced synthetic genes as internal controls in RNA sequencing experiments. (2016) Nature Methods.

PROTOCOL CITATION

Tim Mercer 2020. Using sequins with RNA sequencing.. **protocols.io**
https://protocols.io/view/using-sequins-with-rna-sequencing-bic6kaze

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Hardwick et. al., Spliced synthetic genes as internal controls in RNA sequencing experiments. (2016) Nature Methods.

EXTERNAL LINK

www.sequinstandards.com

KEYWORDS

RNA sequencing, sequins, controls, normalization

CREATED

Jul 08, 2020

LAST MODIFIED

Aug 05, 2020

PROTOCOL INTEGER ID

39038

MATERIALS

NAME	CATALOG #	VENDOR

NAME	CATALOG #	VENDOR
RNA sequins standards		Sequins

DISCLAIMER:

BEFORE STARTING

Install Anaquin Software.

To analyze sequins, we have developed a software toolkit, named *anaquin*, that accepts .FASTQ or .BAM formats, and can be integrated into your RNAseq bioinformatic pipeline. When anaquin rna processes the .FASTQ or .BAM files, it performs two main functions:

(i) Calibrate. The number/fraction of sequin reads in a library can be modulated. For example, anaquin rna can calibrate the number of sequins reads to comprise 1% of the library (using the *–calibrate 0.01* option). This tool is useful for matching dilution between multiple replicates and samples.

(ii) Report. Anaquin rna generates several useful reports, including on library performance (*rna_summary.stats*), quantitative accuracy (*rna_sequins_table.tsv*), and individual sequin performance (*rna_sequins.tsv*).

Anaquin can be downloaded from https://github.com/sequinstandards/Anaquin then run:

```
unzip anaquin_3.14.2.zip
cd anaquin_3.14.2
make
```

Example data.

In this protocol we have used the following example RNAseq libraries that can be downloaded from https://www.sequinstandards.com/resources/

K562_SequinMixA.Rep1.R1.fq.gz
K562_SequinMixA.Rep1.R2.fq.gz

Briefly, total RNA was extracted from K562 cell line with sequins (Mix A) was added. Libraries were prepared using the KAPA Stranded mRNA-SeqTM and sequenced using Illumina HiSeq 2500TM.

Laboratory Steps

1 **Receiving sequins.**
Upon receipt of RNA sequins, first check to ensure they have not thawed during shipment. Please contact us if you have any concerns. Immediately transfer the RNA sequins to frozen storage at -80°C (sequins should not be stored in a -20°C frost-free freezer).

Each tube contains RNA sequins in 10 µL solution, which is typically sufficient for ~100 RNAseq libraries. On first thaw, spin the tube down to collect the contents at the bottom of tube, and prepare smaller single-use aliquots to minimize subsequent freeze-thaw cycles.

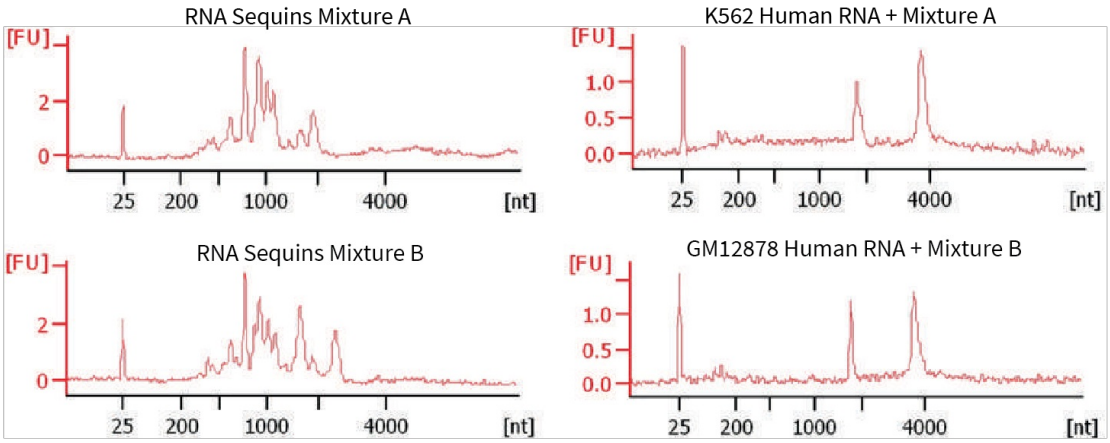

Figure 1. Example traces of RNA sequins using an 2100 BioAnalyzer with the RNA Nano Kit (Agilent Technologies) for (left upper) neat Sequin Mixture A and (left lower) neat Sequins Mixture B. Also shown are example traces for (right upper) K562 with Sequin Mixture A and (right lower) GM12878 with Sequins Mixture B.

2

Preparing sequin stocks.

RNA sequins are provided in solution in nuclease-free water at a concentration of 15 ng/µL. Please use the table below to determine the amount and dilution of sequins that should be added to the sample RNA amount (please note users should dilute sequins even further when using with targeted RNA sequencing applications) :

Sample RNA Amount.	Dilution Volume (ddH20)	Diluted Sequins Concentration.	Sequin Volume to Add to Sample.	Mass of Sequins Added.	Final Sequin Concentration
20ng	740ul	0.2ng/ul	1ul	0.2ng	1%
50ng	290ul	0.5ng/ul	1ul	0.5ng	1%
100ng	140ul	1ng/ul	1ul	1.0ng	1%
500ng	20ul	5ng/ul	1ul	5.0ng	1%
1000ng	6ul	10ng/ul	1ul	10.0ng	1%

Table 1. Guidelines for diluting RNA sequins according to sample RNA amounts (recommended 1% spike-in).

NOTE | RNA sequins are provided in two mixtures (A and B). Each mixture contain the same sequin transcripts, but at different molar ratios, thereby emulating fold-change differences in gene expression and alternative splicing between the two mixtures. Mixture A and B can be added to different samples to allow the detection of these known fold-differences.

3 **Sequencing.**

The library that is generated from the combined RNA sample and sequins is then sequenced according to manufacturer's instructions.

Analysis of .FASTQ libraries (Option 1).

4

Analysis of .FASTQ libraries (option 1).

To directly analyze the sequin from your library .FASTQ files, run the following command:

```
anaquin rna -t 24 -o results --calibrate 0.005 -1 K562_SequinMixA.Rep1.R1.fq.gz \
-2 K562_SequinMixA.Rep1.R2.fq.gz
```

In this example command, we used the *--calibrate 0.005* option to subsample sequin reads to coprise 0.5% of total

reads in the NGS library.

Analysis of .BAM libraries (Option 2).

5

Build index with decoy chromosome.
The sequins can also be aligned to a decoy chromosome (chrQ) that is indexed along with the reference genome assembly.

To build this combined index, the user first concatenates the human genome sequences to the decoy chromosomes into a single file:

```
cat hg38.fa rnasequin_decoychr_2.4.fa >hg38_decoy.fa
```

User then builds an index from these combined files using their alignment tool of choice. For example, using the STAR aligner (Doben et. al., 2013):

```
mkdir /path/to/star_genome_dir
STAR --runMode genomeGenerate \
--genomeDir ./star_genome_dir \
--genomeFastaFiles hg38_decoy.fa
```

6 **Alignment (using STAR).**
We then align the library to the combined index:

```
STAR --runThreadN 8 \
--runMode alignReads \
--genomeDir /path/to/star_genome_dir \
--readFilesIn K562_SequinMixA.Rep1.R1.fq.gz K562_SequinMixA.Rep1.R2.fq.gz \
--readFilesCommand zcat \
--outSAMtype BAM SortedByCoordinate \
--outFileNamePrefix K562_SequinMixA.Rep1
```

7 **Analysis of sequins.**
We finally use anaquin rna to analyse the .BAM alignment files:

```
anaquin rna -t 24 -o results --calibrate 0.005 --combined K562_SequinMixA.Rep1.bam
```

In this example command, we used the *--calibrate 0.005* option to subsample sequin alignments to comprise 0.5% of total alignments from the NGS library.

Output Results.

8

When *anaquin rna* is complete, the following files are generated in the output directory:

Analysis.
anaquin.log – Log files recording the usage and execution of sequin processes.
rna_report.html – Useful visual report describing sequin performance in library. Load with browser.
rna_summary.txt - Summary statistics describing sequins and libraries.
rna_sequins.tsv - Detailed statistics on each individual sequin.
rna_sequins_calibrated.tsv - Detailed statistics on individual sequins following calibration.

Libraries.
rna_sequin_gene_table.tsv – gene-level quantification of sequins.
rna_sequin_isoform_table.tsv – isoform-level quantification of sequins.
*rna_sample_** - Sample alignments/reads (excludes sequins).
*rna_sequins_** - Alignments/reads derived from sequins