# STAR Protocols

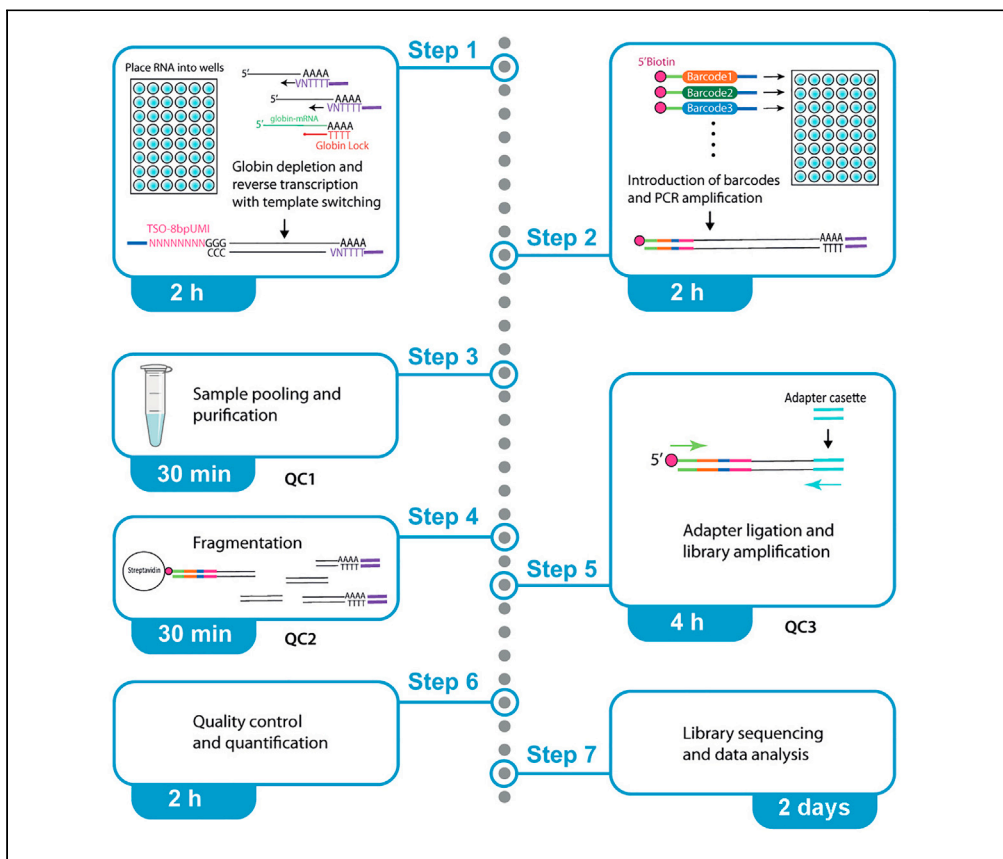## Protocol

# Generation of RNA sequencing libraries for transcriptome analysis of globin-rich tissues of the domestic dog



Sini Ezer, Masahito Yoshihara, Shintaro Katayama, Carsten Daub, Hannes Lohi, Kaarel Krjutskov, Juha Kere

sini.ezer@helsinki.fi (S.E.)
juha.kere@ki.se (J.K.)

## Highlights

transcriptome analysis across tissues of domestic dog, *Canis familiaris*

RNA-seq library preparation for 48 tissue samples in parallel

depletion of abundant globin mRNAs from blood and blood-rich tissues

study of transcription start sites with cDNA reads from 5′end

We have developed a protocol for barcoded cDNA libraries of 48 samples to study gene expression across tissues in the domestic dog, *Canis familiaris,* by modifying the Single-Cell Tagged Reverse Transcription (STRT) protocol. The cDNA reads represent mRNA 5′ ends, enabling the study of transcription start sites (TSS). Our modifications include longer UMIs for molecular counting and Globin-Lock® to deplete globin mRNAs that are abundant in blood and blood-rich tissues dominating all reads.

# STAR Protocols

**Protocol**

# Generation of RNA sequencing libraries for transcriptome analysis of globin-rich tissues of the domestic dog

Sini Ezer,[1,2,9,10,*] Masahito Yoshihara,[3] Shintaro Katayama,[1,2,3] DoGA consortium, Carsten Daub,[3,4] Hannes Lohi,[1,5,6] Kaarel Krjutskov,[7,8] and Juha Kere[1,2,3,*]

[1]Folkhälsan Research Center, 00290 Helsinki, Finland

[2]Stem Cells and Metabolism Research Program, University of Helsinki, 00014 Helsinki, Finland

[3]Department of Biosciences and Nutrition, Karolinska Institutet, 14183 Huddinge, Sweden

[4]Science for Life Laboratory, Stockholm, Sweden

[5]Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

[6]Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland

[7]Competence Centre of Health Technologies, 50411 Tartu, Estonia

[8]Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia

[9]Technical contact

[10]Lead contact

*Correspondence: sini.ezer@helsinki.fi (S.E.), juha.kere@ki.se (J.K.)
https://doi.org/10.1016/j.xpro.2021.100995

## SUMMARY

**We have developed a protocol for barcoded cDNA libraries of 48 samples to study gene expression across tissues in the domestic dog, *Canis familiaris,* by modifying the Single-Cell Tagged Reverse Transcription (STRT) protocol (Islam et al., 2012, 2014). The cDNA reads represent mRNA 5′ ends, enabling the study of transcription start sites (TSS). Our modifications include longer UMIs for molecular counting and Globin-Lock® to deplete globin mRNAs that are abundant in blood and blood-rich tissues dominating all reads.**

## BEFORE YOU BEGIN

### Background and motivation

In the Single-Cell Tagged Reverse Transcription method (STRT-Seq), reads are obtained from 5′ end, which enables study of the transcription start sites. We have modified the method to analyze the RNA extracted from blood, cells, and tissues. Only a small amount, 20–40 ng of template RNA, is required, and thus the method is suitable for studies with limited availability of starting material. Our modifications included longer UMIs, 8 bp instead of 6 bp, which means a sixteen-fold increase in the distinct labels. We have redesigned the primers and streamlined the protocol so that it is possible to complete the library preparation in one day. Second day is needed for quality control. There are several pause points for flexibility.

High content of globin mRNA (gmRNA), which can constitute 50% or even more of the mRNA content in red blood cells, can complicate transcriptome analysis of blood and blood-rich tissues. We attempted to study transcript profile across tissues of domesticated dog *Canis familiaris*. However, after autopsy high content of blood remains within many tissues, such as in bone marrow, liver and spleen; for example, 8%–60% of RNA-seq reads in dog spleen in our preliminary trial were from hemoglobin alpha or beta genes (*data not shown*). We have previously described a fast and effective method, GlobinLock (GL), where gmRNA-specific oligonucleotides are added into initial RNA denaturation step (Krjutškov et al., 2016). Here we describe improved efficiency of the method by

including Locked Nucleic Acids (LNA) in the species-specific oligo design for the dog. The protocol is universal to other species by switching species-specific Globin-Lock® blocking oligonucleotides.

One limiting parameter for use of RNA-Seq has been its high cost, which can be significantly reduced by multiplexing the samples. The size of the library multiplex level in the protocol is 48. Typically, two or more libraries can be processed simultaneously. To prepare multiple libraries per study, the sample layout should be considered carefully as guided by Katayama et al. (2019).

### Animal samples

The dog samples used in the protocol were from pet dogs euthanized by the owner's decision. The Ethical committee of the Faculty of Veterinary Sciences evaluated and accepted the animal proceedings under the ethical permissions ESAVI/7482/04.10.07/2015 and ESAVI/343/04.10.07/2016.

### Workspace setup

To avoid contamination, you should have two separate areas for pre- and post-PCR work with designated consumables and small equipment (mini-centrifuge, vortex and pipettors) in each area. We recommend using two thermocyclers, one in the pre-PCR area for reverse transcription reaction only, and another one for amplifications in post-PCR area.

Refer to "materials and equipment" for the list of buffers and solutions to be prepared before starting the protocol.

### Sample RNA preparations

⏱ Timing: 3 h

1. Analyze the purity and the integrity of the template RNA with TapeStation, Bioanalyzer or Qubit IQ assay. Only samples with RIN>7.5 should be used for optimal data.
2. Dilute the RNA sample to 20 ng/μL and use 2 μL per well (total 40 ng).

⚠ CRITICAL: Keep your RNA sample tubes on ice at all times while making the dilutions.

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Chemicals, peptides, and recombinant proteins | | |
| RiboLock Rnase inhibitor, 40 U/μL | Thermo Fisher Scientific | Cat#EO0382 |
| Phusion DNA polymerase, 2 U/μL | Thermo Fisher Scientific | Cat#F-530S |
| NEBNext End Repair Enzyme | New England Biolabs | Cat#E6050S |
| E.coli DNA ligase | New England Biolabs | Cat#M0205S |
| NEBNext dA-Tailing Reaction Buffer, 10× | New England Biolabs | Cat#B6059S |
| Klenow exo, 5 U/μL | New England Biolabs | Cat#M0212S |
| T4 DNA ligase, 5 U/μL | Thermo Fisher Scientific | Cat#15224-041 |
| SuperScript II Reverse Transcriptase, 200U/μL | Invitrogen | Cat#18064-014 |
| Betaine 5M | Sigma-Aldrich | Cat#B0300-1VL |
| DTT No-Weigh Format | Thermo Fisher Scientific | Cat#A39255 |
| Tris 1 M, pH 8 | Ambion | Cat#AM9855G |
| Ultra Pure Tris 1 M, pH 7.5 | Thermo Fisher Scientific | Cat#15567-027 |
| 2M KCL, RNAase free | Ambion | Cat#AM9640G |
| 1M $MgCl_2$, RNAase free | Ambion | Cat#AM9530G |

*(Continued on next page)*

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| 5M NaCl, RNAase free | Ambion | Cat#AM9760G |
| EDTA 0.5 M, RNAase free | Ambion | Cat#AM9260G |
| NucleoSpin Gel and PCR Clean-up | MACHEREY-NAGEL | Cat#740609.250 |
| Glycerol | Sigma-Aldrich | Cat#G5516-100ML |
| Triton X-100 10% | Sigma-Aldrich | Cat#93443-100ML |
| Tween-20 | Fisher Scientific | Cat#BP337-100 |
| Nuclease-Free Water | Ambion | Cat#AM9937 |
| Critical commercial assays | | |
| ERCC RNA Spike-in Mix | Ambion | Cat#4456740 |
| KAPA Library Quantification Kit | Roche | Cat#KK4824 |
| NucleoSpin Gel and PCR Clean-up | Macherey-Nagel | Cat#740609.250 |
| Dynabeads MyOne C1 Streptavidin | Invitrogen | Cat#65001 |
| Agencourt AMPure XP | Beckman Coulter | Cat#A63881 |
| NextSeq 500/550 High Output v2 kit | Illumina | Cat#FC-404-2005 |
| Oligonucleotides | | |
| Primers for library preparation, see Table 1 | This paper | N/A |
| Sequences for barcode primers, see Table 2 | This paper | N/A |
| Software and algorithms | | |
| STRT2 pipeline b3e589c | This paper | https://github.com/my0916/STRT2 |
| Picard v2.23.4 | https://github.com/broadinstitute/picard | http://broadinstitute.github.io/picard/ |
| HISAT2 v2.2.1 | Kim et al. (2019) | https://daehwankimlab.github.io/hisat2/ |
| SAMtools v1.10 | Li et al. (2009) | http://www.htslib.org/ |
| BEDtools v2.29.2 | Quinlan and Hall (2010) | http://bedtools.readthedocs.io/ |
| R v4.0.0 | R core Team (2020) | https://www.r-project.org/ |
| R Package: ggplot2 v 3.3.3 | Wickham and Grolemund (2016) | https://ggplot2.tidyverse.org/ |
| featureCounts v2.0.0 | Liao et al. (2014) | http://subread.sourceforge.net/ |
| StringTie v2.1.4 | Pertea et al. (2015) | https://ccb.jhu.edu/software/stringtie/ |
| FastQC v0.11.8 | https://github.com/s-andrews/FastQC | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| MultiQC v1.9 | Ewels et al. (2016) | https://multiqc.info/ |
| Other | | |
| Thermocycler with a heated lid | N/A | N/A |
| Focused ultrasonicator for DNA fragmentation | Covaris or compatible | N/A |
| Magnetic stand for 1.5mLtubes | N/A | N/A |
| Thermal Mixer (Thermal Shaker) for 1.5 mL tubes | N/A | N/A |
| Bench Top Centrifuge for 1.5/2 mL tubes | N/A | N/A |
| VersiPlates, 96-well Low Profile PCR Strip Tube Plate | Thermo Scientific | Cat#AB-1800 |
| VersiCap Mat, Domed Cap Stripes | Thermo Scientific | Cat#AB-1810 |
| DNA LoBind Tubes 1.5 mL | Eppendorf | Cat#30108051 |
| 8-channel Multipipette | N/A | N/A |
| TapeStation (for quality control) | Agilent | N/A |
| D1000 ScreenTape | Agilent | Cat#5067- 5582 |
| D1000 Reagents | Agilent | Cat#5067- 5583 |

## MATERIALS AND EQUIPMENT

⚠ CRITICAL: Prepare all the buffers and solutions in the pre-PCR area, preferably in a PCR workstation, and use nuclease-free water.

### ERCC spike-in control dilution (for step 6)

⊙ Timing: 30 min

**Table 1. Primers for library preparation**

| Primer | Sequence | NOTE |
|---|---|---|
| GLα1 (canine globin alpha 1) | T+TTTTG+CAGCCCA+CTCAGACTTT+ATTCAAA+CATC-3Phos | +N = LNA nucleotide |
| GLα2 (canine globin alpha 2) | T+TTTTGGGC+GCAGGAGGC+GGCCCA+G-3Phos | +N = LNA nucleotide |
| GLβ (canine globin beta) | T+TTTT+GAGCTGA+AGGTTCT+TTATTAGG+CAGAAG-3Phos | +N = LNA nucleotide |
| TSO 8 UMI Biotin | Biotin-ACGACGCTCTTCCGATCTNNNNNNNNrGrGrG | |
| T30-VN GL | TTAAGCAGTGGTATCAACGCAGAGTCGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN | |
| universal primer PCR1c | AAGCAGTGGTATCAACGCAGAGT | |
| adapter1 | CAAGCAGAAGACGGCATACGAGATT | |
| adapter2 | Phos-ATCTCGTATGCCGTCTTCTGCTTG | |
| PCR-2cF | CAAGCAGAAGACGGCATACGAGATT | |
| PCR-2cR | AATGATACGGCGACCACCGAGATCTACACCACGTCTGAACTCCAGTCAC | |
| STRT2seq | CGAGATCTACACCACGTCTGAACTCCAGTCAC | |

ERCC RNA Spike-in Mix requires dilution. To avoid repeated thaw and freeze cycles, make aliquots as follows.

- Create ten 1:10 dilutions and then 1:100 dilutions as working stocks.
- Make aliquots of 1.3 μL from the 1:100 solution and store at −70°C (can be stored up to 12 months).
- Thaw an aliquot on ice and take 1 μL for each library master mix.

### Barcode primers (for step 10)

⏱ Timing: 2 h

- Dissolve each Barcode primer (Table 2) in nuclease-free water to make 100 μM stocks (store at −20°C. Stable up to 12 months).
- Prepare 48 well plate with 10 μM barcoded primer solutions in numerical order. Store at −20°C (stable up to 12 months).
- Thaw for 2–3 h at +4°C before use.

### Preparation of double-stranded adapter cassette (for step 44)

⏱ Timing: 1 h

**Adapter cassette 20 μM**

| Reagent | Final concentration | Amount |
|---|---|---|
| Adapter-1 ssDNA (100 μM) | 20 μM | 20 μL |
| Adapter-2 ssDNA (100 μM) | 20 μM | 20 μL |
| KCl (2M) | 50 mM | 2.5 μL |
| Nuclease-Free Water | n/a | 57.5 μL |
| Total | n/a | 100 μL |

- Mix the reagents in a 1.5 mL tube.
- Heat 5 min at 72°C in a heat block.
- Turn the heat block off and let the heat block cool slowly to 20°C–22°C.
- Store the solution at −20°C (stable up to 12 months).
- Thaw on ice before each use.

  *Note:* We recommend using a heat block rather than a water-bath, to allow gradual cooling and annealing to take around 1 h.

**Table 2. Barcode oligos**

| Oligo name | Barcode | 5' modification | Sequence |
|---|---|---|---|
| Barcode 1 | AGTTCA | Biotin | CACGTCTGAACTCCAGTCAC**AGTTCA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 2 | CAGATC | Biotin | CACGTCTGAACTCCAGTCAC**CAGATC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 3 | CGATGT | Biotin | CACGTCTGAACTCCAGTCAC**CGATGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 4 | TTAGGC | Biotin | CACGTCTGAACTCCAGTCAC**TTAGGC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 5 | ATCACG | Biotin | CACGTCTGAACTCCAGTCAC**ATCACG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 6 | TGACCA | Biotin | CACGTCTGAACTCCAGTCAC**TGACCA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 7 | ACATGT | Biotin | CACGTCTGAACTCCAGTCAC**ACATGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 8 | GCCAAT | Biotin | CACGTCTGAACTCCAGTCAC**GCCAAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 9 | CATGAT | Biotin | CACGTCTGAACTCCAGTCAC**CATGAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 10 | TATTGT | Biotin | CACGTCTGAACTCCAGTCAC**TATTGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 11 | AAAGTT | Biotin | CACGTCTGAACTCCAGTCAC**AAAGTT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 12 | TCTACC | Biotin | CACGTCTGAACTCCAGTCAC**TCTACC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 13 | TTGGAC | Biotin | CACGTCTGAACTCCAGTCAC**TTGGAC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 14 | CAAAGT | Biotin | CACGTCTGAACTCCAGTCAC**CAAAGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 15 | ATGCTT | Biotin | CACGTCTGAACTCCAGTCAC**ATGCTT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 16 | GTGGTA | Biotin | CACGTCTGAACTCCAGTCAC**GTGGTA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 17 | GCAGGA | Biotin | CACGTCTGAACTCCAGTCAC**GCAGGA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 18 | GGACAT | Biotin | CACGTCTGAACTCCAGTCAC**GGACAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 19 | ATTCCA | Biotin | CACGTCTGAACTCCAGTCAC**ATTCCA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 20 | AGTTAC | Biotin | CACGTCTGAACTCCAGTCAC**AGTTAC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 21 | TGAAGC | Biotin | CACGTCTGAACTCCAGTCAC**TGAAGC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 22 | TAGCAT | Biotin | CACGTCTGAACTCCAGTCAC**TAGCAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 23 | GTTGCC | Biotin | CACGTCTGAACTCCAGTCAC**GTTGCC**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 24 | ACGTTG | Biotin | CACGTCTGAACTCCAGTCAC**ACGTTG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 25 | TAAGGG | Biotin | CACGTCTGAACTCCAGTCAC**TAAGGG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 26 | GCCTAG | Biotin | CACGTCTGAACTCCAGTCAC**GCCTAG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 27 | CTCGCA | Biotin | CACGTCTGAACTCCAGTCAC**CTCGCA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 28 | CCGAAT | Biotin | CACGTCTGAACTCCAGTCAC**CCGAAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 29 | GGGTTT | Biotin | CACGTCTGAACTCCAGTCAC**GGGTTT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 30 | GTAATG | Biotin | CACGTCTGAACTCCAGTCAC**GTAATG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 31 | TAGAGA | Biotin | CACGTCTGAACTCCAGTCAC**TAGAGA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 32 | AGATGG | Biotin | CACGTCTGAACTCCAGTCAC**AGATGG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 33 | ATCTCT | Biotin | CACGTCTGAACTCCAGTCAC**ATCTCT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 34 | CGTATT | Biotin | CACGTCTGAACTCCAGTCAC**CGTATT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 35 | ACTTAT | Biotin | CACGTCTGAACTCCAGTCAC**ACTTAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 36 | GATCTT | Biotin | CACGTCTGAACTCCAGTCAC**GATCTT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 37 | GCTGTG | Biotin | CACGTCTGAACTCCAGTCAC**GCTGTG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 38 | ACAATA | Biotin | CACGTCTGAACTCCAGTCAC**ACAATA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 39 | TTCATA | Biotin | CACGTCTGAACTCCAGTCAC**TTCATA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 40 | TTAACT | Biotin | CACGTCTGAACTCCAGTCAC**TTAACT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 41 | ATACAG | Biotin | CACGTCTGAACTCCAGTCAC**ATACAG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 42 | AGTCGT | Biotin | CACGTCTGAACTCCAGTCAC**AGTCGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 43 | CTGTGT | Biotin | CACGTCTGAACTCCAGTCAC**CTGTGT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 44 | CCTAGA | Biotin | CACGTCTGAACTCCAGTCAC**CCTAGA**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 45 | CCATCT | Biotin | CACGTCTGAACTCCAGTCAC**CCATCT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 46 | GACACT | Biotin | CACGTCTGAACTCCAGTCAC**GACACT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 47 | TGGATG | Biotin | CACGTCTGAACTCCAGTCAC**TGGATG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| Barcode 48 | CTCCAT | Biotin | CACGTCTGAACTCCAGTCAC**CTCCAT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT |

The barcode sequence within each primer is shown in bold.

**Buffer EB**

| Reagent | Final concentration | Amount |
| --- | --- | --- |
| Tris-HCl (pH 8), (1 M) | 5 mM | 0.25 mL |
| Tween-20 (10%) | 0.02% | 0.1 mL |
| Nuclease-Free water | n/a | 49.65 mL |
| Total | n/a | 50 mL |

Store at 20°C–22°C up to 12 months.

**Buffer 2×BP**

| Reagent | Final concentration | Amount |
| --- | --- | --- |
| Tris-HCl (pH 7.5), (1 M) | 10 mM | 0.5 mL |
| EDTA (0.5 M) | 10 mM | 1 mL |
| NaCl (5 M) | 2 M | 20 mL |
| Tween-20 (10%) | 0.01% | 0.05 mL |
| Nuclease-Free water | n/a | 28.45 mL |
| Total | n/a | 50 mL |

Store at 20°C–22°C up to 12 months.

*Note:* All the reagents used for buffers A-GL, B-GL, C and buffer C concentrate should be nuclease free.

**Buffer A-GL**

| Reagent | Final concentration in RT reaction (6 μL) | Amount for 1 rxn | Master mix for 1 library |
| --- | --- | --- | --- |
| Nuclease free water | n/a | 1.03 μL | 56.65 μL |
| dNTP mix (10 mM each) | 1 mM | 0.6 μL | 33 μL |
| KCl (2 M) | 75 mM | 0.22 μL | 12.1 μL |
| Triton X-100 (10%) | 0.05% | 0.03 μL | 1.65 μL |
| Tris-HCl (pH 7.5), (1 M) | 5 mM | 0.03 μL | 1.65 μL |
| Reagent | Final concentration in GL (3 μL) | Amount for 1 rxn | Master mix for 1 library |
| primer GLα1 (1 mM) | 10 μM | 0.03 μL | 1.65 μL |
| primer GLα2 (1 mM) | 10 μM | 0.03 μL | 1.65 μL |
| primer GLβ (1 mM) | 10 μM | 0.03 μL | 1.65 μL |
| Total | n/a | 2.0 μL | 110.0 μL |

Store at −20°C up to 6 months.

**Buffer B-GL**

| Reagent | Final concentration in RT (6 μL) | Amount for 1 rxn | Master mix for 1 library |
| --- | --- | --- | --- |
| Betaine (5 M) | 1 M | 1.113 μL | 66.78 μL |
| Tris-HCl (pH 8.5), (1 M) | 20 mM | 0.12 μL | 7.2 μL |
| primer T30-VN GL (100 μM) | 0.2 μM | 0.012 μL | 0.72 μL |
| DTT (500 mM) | 10 mM | 0.12 μL | 7.2 μL |
| primer TSO-8UMI (100 μM) | 1.5 μM | 0.09 μL | 5.4 μL |
| MgCl$_2$ (1 M) | 7.5 mM | 0.045 μL | 2.7 μL |
| Total | n/a | 1.5 μL | 90.0 μL |

Store at −20°C up to 6 months.

**Buffer C**

| Reagent | Final concentration in RT (6 µL) | Amount for 1 rxn | Master mix for 1 library |
|---|---|---|---|
| Glycerol | | 0.2724 µL | 13.1 µL |
| SuperScript II (200 U/µL) | 3.3 U/µL | 0.1 µL | 5.9 µL |
| RiboLOCK (40 U/µL) | 0.7 U/µL | 0.1 µL | 6.3 µL |
| DTT (500 mM) | 5 mM | 0.06 µL | 3.6 µL |
| C buffer concentrate (see below) | n/a | 0.0176 µL | 1.1 µL |
| Total | n/a | 0.5 µL | 30.0 µL |

Store at −20°C up to 6 months.

> *Note:* Buffer C is very viscous, and thus larger excess volume is calculated for the master mixes of Buffer B and C. To pipette viscous solutions accurately pause longer during aspiration, as well as dispensing.

**C Buffer concentrate stock**

| Reagent | Final concentration | Amount |
|---|---|---|
| NaCl (5 M) | 100 mM | 100 µL |
| Tris-HCl (pH 7.5), (1 M) | 5 mM | 25 µL |
| Triton X-100 (10%) | 0.10% | 50 µL |
| EDTA (0.5 M) | 0.1 mM | 1 µL |
| Total | n/a | 176 µL |

Store at −20°C up to 12 months.

> *Note:* Final concentration of C buffer concentrate is calculated for the RT-reaction.

**PCR-2 primer mix (for step 50)**

| Reagent | Final concentration | Amount |
|---|---|---|
| primer PCR-2cF (100 µM) | 10 µM | 10 µL |
| primer PCR-2cR (100 µM) | 10 µM | 10 µL |
| Nuclease free water | n/a | 80 µL |
| Total | n/a | 100 µL |

Store at −20°C up to 12 months.

## STEP-BY-STEP METHOD DETAILS

Figure 1 shows the schematic overview of the method.

**Globin-Lock® and reverse transcription with template switching**

🕐 Timing: 2 h

1. Before starting, get your cycler ready for the STRT-GL reaction at 80°C.
2. Mix **2 µL RNA** 20 ng/µL (47 samples) and **2 µL Buffer A-GL Master mix** on each sample well on VersiPlate.
3. For negative control, mix **2 µL H$_2$O** with **2 µL buffer A-GL Master mix**.
4. Seal the plate and spin drops down (1 min at 1500 × *g*).
5. Proceed to the RNA denaturation step and leave the plate at 40°C on hold, until you are ready with step 6.
6. Prepare RT mixture of **90 µL Buffer B-GL Master mix** and **30 µL Buffer C Master mix** per one library (48 wells) and add **1 µL Spike-In** (1:100 dilution) per one library.
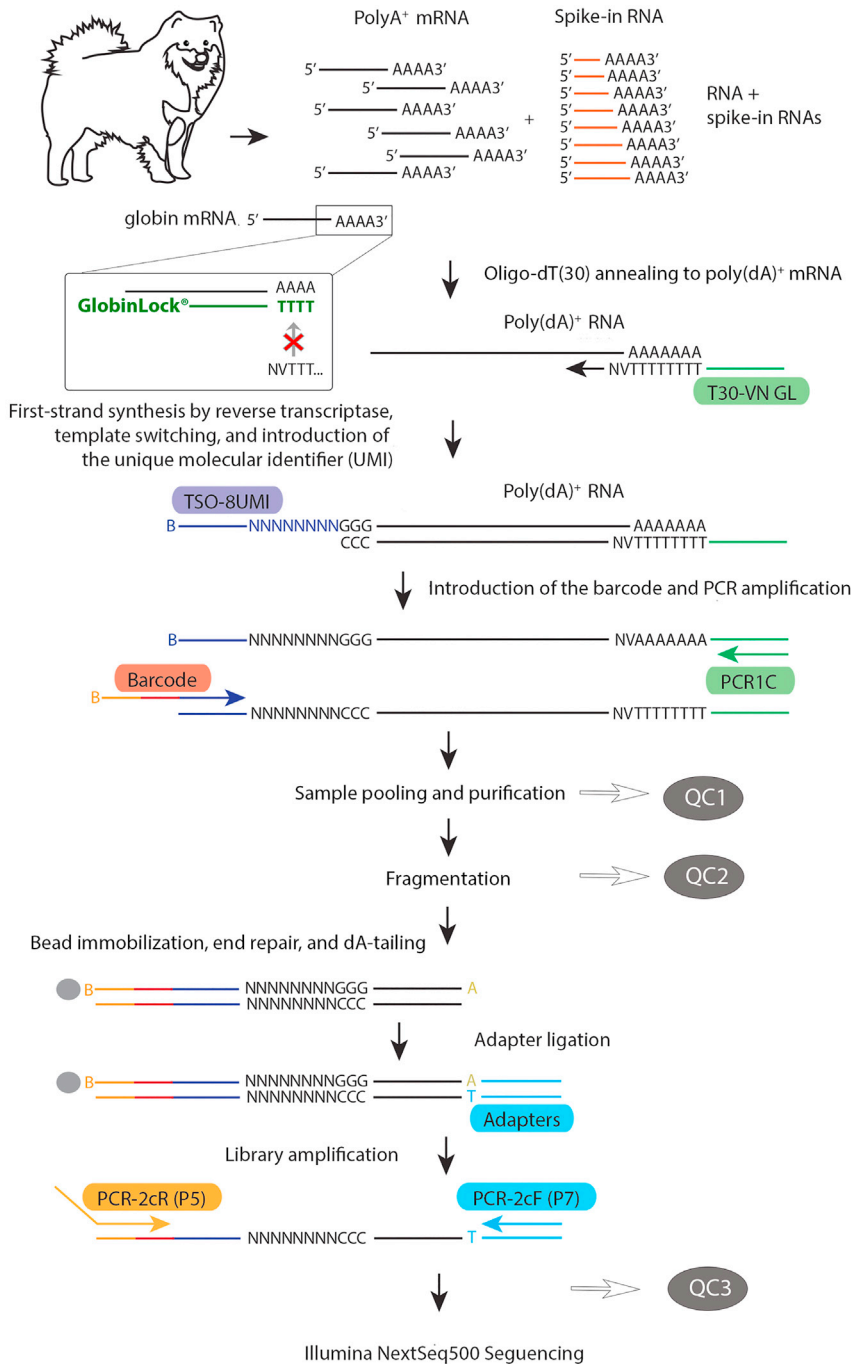
**Figure 1. Overview of the library preparation with the required primers**
Template RNA is mixed with ERCC Spike-in RNA, which is added equally to each sample for normalization, and with globin-LNA primers, which hybridize to the sequence adjacent to poly-A$^+$ tail of globin mRNA thus blocking its' transcription. First-strand cDNA synthesis occurs using an oligo(dT) primer with the addition of 3-6 cytosines. Oligo TSO-8UMI promotes template switching, which introduces UMIs into the cDNA. The cDNA is amplified, and the well specific barcodes are introduced at the 5'end. The samples are then pooled and ligated to an adapter cassette. The 5'end is further amplified for sequencing. QC1 to QC3 show the steps where a sample is taken for quality control.

7. Add **2 μL** RT mixture to each well on 40°C. Total volume in each well is now 6 μL. Mix by pipetting up and down several times.
8. Continue with **STRT-GL** reaction.

**STRT-GL reaction conditions**

| Steps | Temperature | Time |
|---|---|---|
| Warm-up/hold | 80°C | hold |
| RNA Denaturation | 80°C | 1 min |
| Globin-Lock Hybridization | 65°C | 2 min |
| Loading | 40°C | hold |
| RT Reaction | 40°C | 60 min |
| Inactivation | 85°C | 5 min |
| Hold | 4°C | forever |

9. Spin down (1 min at 1500 × *g*).

   *Note:* To avoid evaporation of the small reaction volume, we recommend using low-profile VersiPlates. However, you need to check the compatibility of the plates with your thermocycler. We are using Bio-Rad T100 Thermal Cycler with PCR plate Pressure Pad to ensure tight sealing of the lid.

### Introduction of barcodes and full transcript amplification

⏱ Timing: 2 h

10. Keep the reaction plate on ice and add **2 μL** well-specific barcoded primer (10 μM) to each sample.
11. Make the following PCR master mix.

| Reagent | Amount for 1 rxn | Master mix for one library |
|---|---|---|
| Nuclease-Free Water | 7.4 μL | 407 μL |
| Phusion GC buffer, 5× | 4 μL | 220 μL |
| DMSO (100%) | 0.6 μL | 33 μL |
| dNTP mix (each 10 mM) | 0.4 μL | 22 μL |
| Phusion DNA polymerase (2 U/μL) | 0.4 μL | 22 μL |
| Universal primer PCR1c (100 μM) | 0.2 μL | 11 μL |
| **Total volume** | **13 μL** | **715 μL** |

12. Pipette **13 μL** PCR Master mix to each well in the reaction plate and mix by pipetting up and down several times.
13. Seal the the plate and move to post-PCR area.
14. Start the PCR program **STRT-P1.**

**STRT-P1 PCR cycling conditions**

| Steps | Temperature | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 98°C | 1 min | 1 |
| Denaturation | 98°C | 10 s | 10 cycles |
| Annealing | 62°C | 30 s | |
| Extension | 72°C | 5 min | |
| Hold | 4°C | forever | |

**Sample pooling and purification**

⏱ **Timing: 30 min**

15. Spin the plate (1 min at 1500 × *g*). Pool the samples from all the wells into one 1.5 mL LoBind tube.
16. Purify the content with PCR Clean-up column (NucleoSpin Gel and PCR Clean-up/Macherey-Nagel) according to manufacturer's instructions. https://www.mn-net.com/media/pdf/02/1a/74/Instruction-NucleoSpin-Gel-and-PCR-Clean-up.pdf

    *Alternatives:* You can also use equivalent kit for clean-up from other suppliers; however, we have not tested the impact of alternatives on protocol performance.

17. Elute with **32 μL** elution buffer from the kit into a clean 1.5 mL LoBind tube.
18. Take **2 μL** aside for TapeStation (HS D1000) for first quality control (**sample QC-1**).
    a. **30 μL** is left for the next step.
    b. Keep the **QC-1** sample on ice or at −20°C until you are ready with your **QC-2** sample (step 26) and analyze them simultaneously.

    ⧗⧗ **Pause point:** The sample can be stored at −20°C up to 1 month.

**Fragmentation of full-length cDNA**

⏱ **Timing: 30 min**

19. Mix AMPure XP beads well. (You should have homogenous brown solution and no bead precipitates. It is recommended to take an aliquot of 200–500 μL into a 2 mL tube and mix by vortexing before each use.)
20. Add **24 μL** AMPure XP beads to the **30 μL** cDNA tube (from step 18). Mix well by vortexing and keep 10 min at 20°C–22°C.
21. Place the tube on the magnetic stand for at least 3 min or until the supernatant is clear and discard the supernatant.
22. Remove the tube from the magnetic rack and re-suspend the beads in **100 μL** EB. Wait for at least 1 min.
23. Capture the beads on a magnet and transfer the supernatant (**100 μL**, containing the cDNA) into a clean 1.5 mL LoBind tube.
24. Shear the amplified cDNA with Covaris Focused ultrasonicator (or compatible), using 130 μL microtubes. Set the parameters to produce 300 bp fragments according to the manufacturer's protocol. https://www.covaris.com/protocols/?sku=500295
25. Purify and concentrate the fragmented product with PCR Clean-up column (NucleoSpin Gel and PCR Clean-up). Elute with **32 μL** elution buffer from the kit into a clean 1.5 mL LoBind tube.
26. Take **2 μL** aside for TapeStation for the second quality control step (**sample QC-2**). **30 μL** is left for the next step.
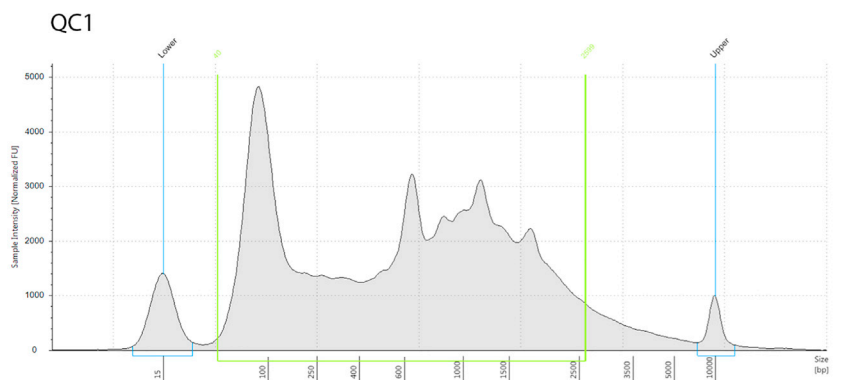
    ⧗⧗ **Pause point:** The sample can be stored at −20°C for up to 1 month.

27. Analyze the **QC-1** (from step 18) and **QC-2** samples (from step 26) with Tape Station before proceeding with the protocol. See Figure 2 for expected analysis results. Typical concentrations are:
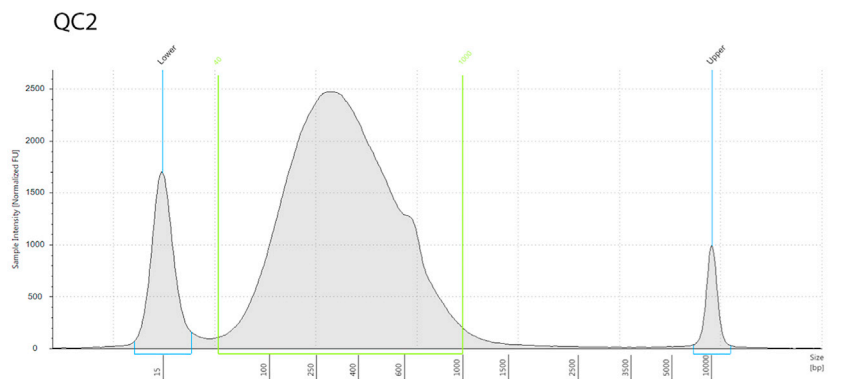
QC-1: 5–50 ng/μL

QC-2: 2–20 ng/μL

If the analysis results are as expected, proceed to the next step. Troubleshooting 1 and 2.

## QC1
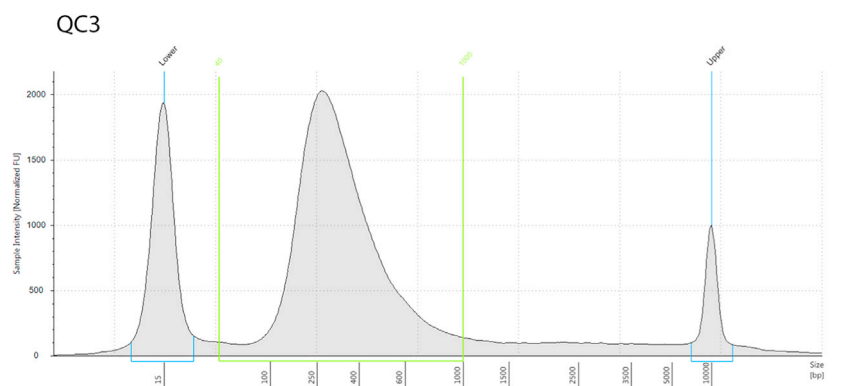


**Region Table**

| From [bp] | To [bp] | Average Size [bp] | Conc. [pg/µl] | Region Molarity [pmol/l] | % of Total | Region Comment | Color |
|-----------|---------|-------------------|---------------|--------------------------|------------|----------------|-------|
| 40 | 2599 | 779 | 11800 | 76900 | 93.22 | | 🟩 |

## QC2



**Region Table**

| From [bp] | To [bp] | Average Size [bp] | Conc. [pg/µl] | Region Molarity [pmol/l] | % of Total | Region Comment | Color |
|-----------|---------|-------------------|---------------|--------------------------|------------|----------------|-------|
| 40 | 1000 | 356 | 4860 | 31000 | 96.06 | | 🟩 |

## QC3



**Region Table**

| From [bp] | To [bp] | Average Size [bp] | Conc. [pg/µl] | Region Molarity [pmol/l] | % of Total | Region Comment | Color |
|-----------|---------|-------------------|---------------|--------------------------|------------|----------------|-------|
| 40 | 1000 | 339 | 2540 | 15300 | 84.56 | | 🟩 |

**Figure 2. Library quality control with TapeStation**

In QC1, after amplification of the full-sized cDNA, you should see variation of transcript sizes (from about 200 bp to >2000 bp). In QC2, the cDNA has been fragmented to size around 300 bp. In QC3 after further amplification, size selection and purification, sharper peak around 300 bp should be visible.

## End-repair, adapter ligation and 5′end amplification

🕐 Timing: 4 h

⚠ CRITICAL: In the following enzymatic steps, incubations are performed with cDNA bound to beads; gentle mixing of the reaction tube is required to avoid sinking of the beads to the bottom of the tube. We have used Eppendorf ThermoMixer C, but equivalent thermal mixer from other suppliers can also be used.

*Alternatives:* If you do not have a thermal mixer, the mixing can be done manually by pipetting at 10 min intervals to disperse the beads.

28. Wash Dynabeads MyOne C1 Streptavidin beads as follows:
    a. Take **12 μL** MyOne C1 Streptavidin beads in a 1.5 mL LoBind tube.
    b. Capture the beads using a magnet, and withdraw the supernatant from the tube.
    c. Remove the tube from the magnetic rack and re-suspend the beads in **200 μL** 2×BP buffer. Capture the beads and remove the supernatant.
    d. Repeat step 28c. once more.
    e. Re-suspend the beads in **30 μL** 2×BP buffer.
29. Mix the washed beads (**30 μL**) and the fragmentized cDNA (**30 μL**, from step 26) in a 1.5 mL LoBind tube. Final concentration of the binding buffer will be now 1xBP.
30. Incubate 20 min on 500 rpm in thermal mixer at 20°C–22°C.
31. Place the tube on a magnetic rack until the beads clearly separate from the liquid (1–3 min). Discard the supernatant and remove the tube from the magnetic rack.
32. Add **200 μL** of EB buffer. Re-suspend the beads thoroughly by pipetting, and then place the tube again on a magnetic stand (1–3 min).
33. Discard the supernatant and re-suspend the beads in **16 μL** buffer EB.
34. Add the following reagents:

| Reagent | Amount |
| --- | --- |
| Beads containing cDNA in EB (from step 33) | 16 μL |
| NEBNext End Repair Buffer, 10× | 2 μL |
| NEBNext End Repair Enzyme Mix | 1 μL |
| *E.coli* ligase | 1 μL |
| Total | 20 μL |

35. Incubate 20 min in a thermal shaker on 500 rpm at 20°C–22°C.
36. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant from the tube. Remove the tube from the magnetic rack and re-suspend the beads in **200 μL** EB.
37. Repeat step 36 once more.
38. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant.
39. Re-suspend in **17 μL** EB. Add the following reagents:

| Reagent | Amount |
| --- | --- |
| Beads containing cDNA in EB (from step 39) | 17 μL |
| NEBNext dA tailing buffer, 10× | 2 μL |
| Klenow exo- (5 U/μL) | 1 μL |
| Total | 20 μL |

40. Incubate 20 min in a thermal shaker on 500 rpm at 37°C.

41. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant from the tube. Re-move the tube from the magnetic rack and re-suspend the beads in **200 µL** EB.
42. Repeat step 40 once more.
43. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant.
44. Re-suspend in **18 µL** EB. Add the following reagents:

| Reagent | Amount |
| --- | --- |
| Beads containing cDNA in EB (from step 44) | 18 µL |
| T4 DNA Ligase buffer (10×) | 3 µL |
| Adapter cassette (20 µM) | 1 µL |
| PEG-4000 (50%) | 6 µL |
| T4 DNA ligase (5 U/µL) | 2 µL |
| Total | 30 µL |

45. Mix gently by pipetting. Do not vortex. Incubate 60 min at 20°C–22°C and mix every 15 min during the incubation by pipetting gently up and down.
46. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant from the tube. Remove the tube from the magnetic rack and re-suspend the beads in **200 µL** EB.
47. Repeat step 46 twice more.
48. Place the tube on a magnetic rack for 1–3 min and withdraw the supernatant.
49. Re-suspend the beads in **20 µL** EB.
50. Transfer the solution with the beads into a 0.2 mL PCR tube. Add the following reagents and start the **KIN-P2** program.

| Reagent | Amount |
| --- | --- |
| Beads containing cDNA in EB (from step 49) | 20 µL |
| Phusion Hot master mix, 2× | 25 µL |
| PCR-2 primer mix (10 µM each) | 5 µL |
| Total | 50 µL |

| KIN-P2 PCR cycling conditions | | | |
| --- | --- | --- | --- |
| Steps | Temperature | Time | Cycles |
| Initial Denaturation | 98°C | 1 min | 1 |
| Denaturation | 98°C | 10 s | 10 cycles |
| Annealing | 62°C | 20 s | |
| Extension | 72°C | 20 s | |
| Final Extension | 72°C | 1 min | 1 |
| Hold | 4°C | forever | |

51. Transfer the reaction into a 1.5 mL LoBind tube and capture the streptavidin beads. Collect the supernatant, **50 µL** in volume, which now contains the PCR product, and transfer it into a clean tube.
52. Add **45 µL** AMPureXP beads. Mix well by vortex.
53. Incubate 10 min at 20°C–22°C and place the tube on the magnetic stand for at least 3 min or until the supernatant is clear.
54. Discard the supernatant. Spin 10 s on a mini-centrifuge (or 11,000 × *g* on a table top centrifuge) to collect drops.

55. Place back on magnet and remove all the traces of supernatant.
56. Resuspend the beads in **20 μL** EB.
57. Wait for 1–3 min and place on magnet.
58. Transfer the clear supernatant (ready library) to a clean 1.5 mL LoBind tube.
59. Take **2 μL** for TapeStation for final quality control (**sample QC-3**).
60. Store the ready library at −20°C.

⏸ **Pause point:** The ready library can be stored at −20°C for up to 6 month.

## Library quality control and concentration

⊙ Timing: 2 h

61. Analyze the **QC-3** sample (from step 59) with TapeStation D1000 ScreenTape system. See Figure 2 for expected analysis results.
62. Determine the concentration of the library with the KAPA Library Quantification Kit. (https://pim-eservices.roche.com/eLD/api/downloads/ca670ceb-fb38-eb11-0291-005056a71a5d?countryIsoCode=pi).

The typical concentration for the ready library is 2–30 nM. Troubleshooting 3.

⚠ CRITICAL: Accurate molar concentration of the library is required to determine how much to load for sequencing. You can determine the concentration from the QC3 sample with TapeStation, however, we have measured it also using KAPA Library Quantification Kit. If there is a difference between the two measurements, we have relied on the result from the KAPA Kit.

## Sequencing
We have sequenced the libraries with NextSeq 500 using High Output v2 kit, 75 cycles. Sequencing was done with custom primer STRT2seq (Table 1).

Library loading concentration was 2.0 pM and we have added PhiX to 20%.

## Sequence data processing
Timing: typically up to 1 day, depending on the sequencing depth and size of the library

This section describes the bioinformatics pipeline processing the STRT sequencing data from Illumina Base Call Format (.bcl), including demultiplexing, mapping, quality check, and quantification. For a detailed analysis process and settings, the pipeline can be accessed at: https://github.com/my0916/STRT2.

63. Install the STRT2 pipeline from the GitHub repository. Type the following command in the terminal window and hit the enter key.

```
git clone https://github.com/my0916/STRT2.git
```

64. Install the dependencies.
    a. Mandatory software are listed below. The following versions were used by the authors in preparing this protocol, while different versions may work correctly.
        i. Picard (v2.23.4)
        ii. HISAT2 (v2.2.1)

    iii.  SAMtools (v1.10)

    iv.  Bedtools (v2.29.2)

    v.  Subread (v2.0.0)

    vi.  R (v4.0.0)

    vii.  Ruby (v2.6.2)

b. Software required for the optional tools are listed below. The following versions were used by the authors in preparing this protocol, while different versions may work correctly.

    i.  StringTie (v2.1.4)

    ii.  FastQC (v0.11.8)

    iii.  MultiQC (v1.9)

c. Alternatively, conda environment can be created from `condaEnv.yml` with the following commands. Please make sure that Anaconda or Miniconda is installed.

```
conda env create -f condaEnv.yml

conda activate STRT2_env
```

d. For Swedish UPPMAX (https://www.uppmax.uu.se/) users, these software are available through the `module` command in the scripts (`STRT2-UPPMAX.sh`, `STRT2-TFE-UP-PMAX.sh`, and `fastq-fastqc-uppmax.sh`).

65. Prepare the index and dictionary before running the pipeline. Details are described in https://github.com/my0916/STRT2#How-to-build-HISAT2-index. Here we describe the case for the dog genome canFam3.

a. Obtain the genome sequences of reference and ERCC spike-ins.

```
wget http://hgdownload.cse.ucsc.edu/goldenPath/canFam3/bigZips/canFam3.fa.gz

unpigz -c canFam3.fa.gz | ruby -ne '$ok = $_ !~ /^>chrUn_/ if $_ =~ /^>/; puts $_ if $ok'>
canFam3_reference.fasta

wget https://www-s.nist.gov/srmors/certificates/documents/SRM2374_putative_T7_products_
NoPolyA_v2.FASTA

cat SRM2374_putative_T7_products_NoPolyA_v2.FASTA >> canFam3_reference.fasta
```

b. Extract splice sites and exons from a GTF file using the Python scripts implemented in HISAT2 (v2.1.0) (Kim et al., 2019). Here Ensembl transcript map (canFam3.transMapEnsemblV4.gtf.gz) was downloaded from the UCSC Table Browser.

```
unpigz -c canFam3.transMapEnsemblV4.gtf.gz | hisat2_extract_splice_sites.py - | grep -v
^chrUn> canFam3.ss

unpigz -c canFam3.transMapEnsemblV4.gtf.gz | hisat2_extract_exons.py - | grep -v ^chrUn>
canFam3.exon
```

c. Build the HISAT2 index of the reference genome and ERCC Spike-in RNAs with the `hisat2-build` function, which outputs a set of files with the suffix '`.ht2`'.

```
hisat2-build  canFam3_reference.fasta  --ss  canFam3.ss  --exon  canFam3.exon  canFam3_
reference
```

d. Create the sequence dictionary for the reference and Spike-in sequences with Picard (v2.23.4) (http://broadinstitute.github.io/picard/) `CreateSequenceDictionary`.

```
java -jar picard.jar CreateSequenceDictionary R=canFam3_reference.fasta O=canFam3_
reference.dict
```

66. Run the STRT2 pipeline with the following command. Please make sure that barcode sequence with barcode name (Table 2) is prepared as `barcode.txt`.

```
./STRT2.sh -o {output} -g {reference_genome} -a {annotation}\

-b {path/to/BaseCalls} -i {path/to/index} -c {Sequencing_center}\

-r {Run_barcode}
```

Wherein:
`{output}` defines the prefix of output files.
`{reference_genome}` is the reference genome such as canFam3.
`{annotation}` is the gene annotation used for quality check and counting. For canFam3, `ref` (RefSeq) or `ens` (Ensembl) can be chosen. See also the GitHub repository's README.md file for details.
`{path/to/BaseCalls}` is the path to the Illumina basecalls directory.
`{path/to/index}` is the directory and basename of the HISAT2 index for the reference genome.
`{Sequencing_center}` is the name of the sequencing center that produced the reads.
`{Run_barcode}` is the barcode of the run. Prefixed to read names.

> *Note:* In this case, set the parameters '–genome canFam3' and '–annotation ens'. In the default settings, the annotation for quality check and quantification is set to RefSeq.
> Details of the pipeline are as follows.

    a. The raw base call (BCL) files are demultiplexed with Picard `ExtractIlluminaBarcodes` and `IlluminaBasecallsToSam` to generate unaligned BAM files based on the well-specific barcodes (Table 2).

    b. The unaligned BAM files are converted to FASTQ files with Picard `SamToFastq` and aligned using HISAT2 to the dog reference genome (canFam3) and ERCC Spike-in RNAs (SRM 2374) [NIST SRM 2374 Certificate of Analysis, https://www-s.nist.gov/srmors/certificates/2374.pdf, 2013.] with the Ensembl transcript map provided by UCSC (transMapEnsemblV4) as a guide of exon junctions.

    c. The aligned BAM files are merged with the original unaligned BAM files to create a unique molecular identifier (UMI)-annotated BAM files by Picard `MergeBamAlignment`.

    d. The UMI-annotated BAM files corresponding to each sample derived from 4 lanes are merged using Picard `MergeSamFiles`, and potential PCR duplicates are marked with Picard `Markduplicates`.

    e. For quality check, the resulting BAM files per sample are indexed, and then all the qualified reads, reads without potential duplicates, and total mapped reads are counted based on SAM flags using SAMtools (v1.9) (Li et al., 2009). In addition, reads mapped to protein-coding genes and Spike-in RNAs, as well as 5′-ends of them, are counted using BEDtools (v2.29.2) (Quinlan and Hall. 2010). Values of i) $\log_{10}$ of total mapped read counts, ii) mapping rates, iii) $\log_{10}$ of Spike-in RNA read counts, iv) ratios of total mapped read counts versus Spike-in RNA read counts, v) 5′-end capture rates of Spike-in RNAs, and vi) 5′-end capture rates of protein-coding genes are plotted with the R (v4.0.0) (R core Team, 2020) package

ggplot2 (v3.3.3) (Wickham and Grolemund 2016), marking the outlier samples with the barcode numbers. In Figure 2, globin mRNA rates were calculated by dividing the reads mapped to hemoglobin alpha and beta (HBA and HBB) genes by all the qualified reads per sample.

f. For quantification, Subread featureCounts (v2.0.0) (Liao et al., 2014) is used to assign the reads to 5′-end of genes with parameters '-s 1 –largestOverlap –ignoreDup –primary'. Here, uniquely mapped reads within the 5′- UTR or 500 bp upstream of the protein-coding genes and the first 50 bp of Spike-in sequences are counted.

*Optional:* In addition, one can perform TFE (transcript far 5′-end)-based analysis originally developed in Töhönen et al. (2015), by running STRT-TFE.sh or STRT2-TFE-UPPMAX.sh. Here, TFEs are defined as the first exon (5′-end region) of assembled STRT reads mapped to the genome using StringTie (v2.1.4) and assigned with unique IDs. TFE-level reads are also quantified as described above. Details are described in the GitHub repository's TFE-README.md.

*Note:* If TFE-based analysis is planned, add the option '–dta' (downstream-transcriptome-assembly) for STRT2.sh or STRT2-UPPMAX.sh, which is required in the HISAT2 mapping process, that helps the transcript assembly by StringTie.

*Optional:* After running the pipeline above, one can generate fastq files for each sample from the output BAM files in the fastq directory by running fastq-fastQC.sh or fastq-fastQC-UPPMAX.sh. These fastq files (without duplicated reads) can be submitted to public sequence databases. FastQC files are also generated for each fastq file in the fastqc directory. Based on the FastQC results, MultiQC report (MultiQC_report.html) is generated.

## EXPECTED OUTCOMES

The expected results of library quality control stages with TapeStation D1000 ScreenTape system are shown in Figure 2. The typical concentration for the ready library is 2–30 nM.

To study the effect of Globin-Lock®, we designed two libraries with identical layouts, one with the addition of Globin-Lock® step and the other without. We selected several tissues, including those with a high blood content, such as liver and spleen, and in total 40 ng of purified RNA was used for 48-plex libraries. Figure 3. Shows the comparison of the a) relative Globin mRNA amount and b) mapped reads in each library. The amount of globin mRNAs was highest in spleen and in bone marrow. However, the depth of sequencing improved with Globin-Lock® also in other tissues, and especially in tissues with a high amount of globin mRNAs.

## QUANTITATIVE AND STATISTICAL ANALYSIS

Statistics of redundancy, mapped reads, and mapped rate for an example library is shown in Figure 4. After sequencing, expected redundancy for a successful library is ∼1.5, mapped reads should be ≥ 1M, and mapped rate 60%–80% in the latest dog genome.

The library contained 46 RNA samples from different dog tissues and two non-template controls, and the sequences were mapped to Canine genome. Please note that spike-in reads are included as mapped reads in the pipeline.

## LIMITATIONS

High quality of RNA is crucial for optimal results, and in addition to RNA integrity, impurities such as salt residues can affect the efficiency of reverse transcription. From some tissues, such as gut and adipocytes, it can be extremely difficult to extract RNA in good quality. The quality and amount of the starting material also affects the final yield of the cDNA library. The number of PCR cycles
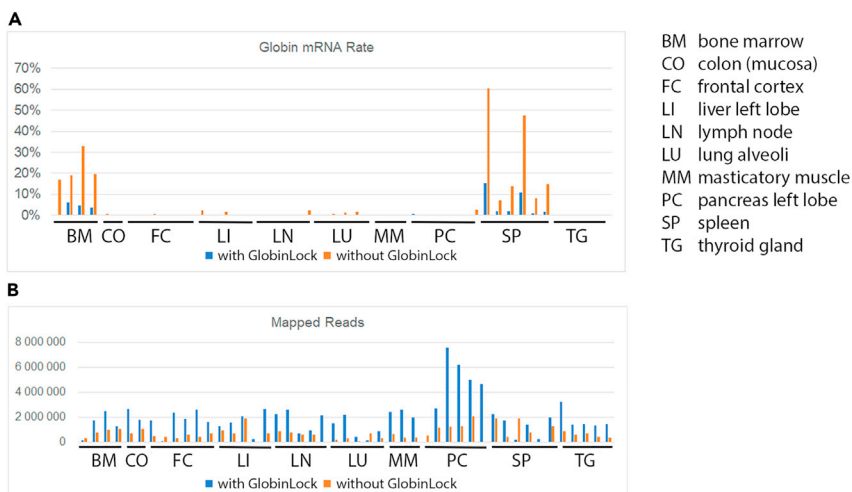
**Figure 3. Effect of Globin-Lock on mapped reads**
Two libraries with identical layouts, one with the addition of Globin-Lock® step and the other without were compared.
(A) Relative amount of globin mRNA in the sample RNA.
(B) Mapped reads in the sample RNA. For each tissue, several replicates are shown from different dogs; the number of replicates were two for colon, three for masticatory muscle, four for bone marrow, five for liver, lymph node, lung and thyroid gland, and six for frontal cortex, pancreas, and spleen. The depth of sequencing improved with Globin-Lock® in many tissues, especially in tissues with a high amount of globin mRNAs, such as spleen and bone marrow.

given in the protocol usually gives enough material for sequencing, however, the number of cycles needs to be calibrated for each application to find the least number of cycles required.

One of the advantages of the protocol is that it allows identification of transcription start sites. However, since the reads are obtained only from the 5′ end of each cDNA, it is not suitable for analysis of splice isoforms.

## TROUBLESHOOTING
### Problem 1
The amount of cDNA products in QC1 is low (step 27).

### Potential solution
Check the quality of your RNA samples. It should not contain genomic DNA, and absorbance ratios 260/280 nm and 260/230 nm should both be close to 2.
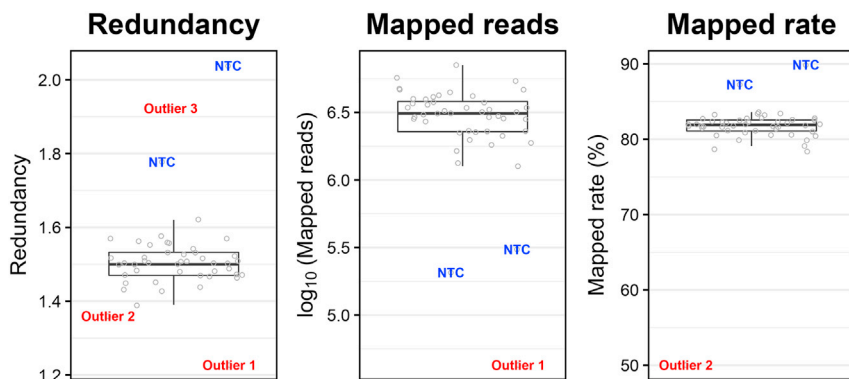


**Figure 4. Output from the sequence analysis**
Redundancy, mapped reads, and mapped rate from a successful library containing 46 RNA samples of different dog tissues and two non-template controls (NTC).

### Problem 2
Dominant peak at around 100 bp in the QC1 (step 27).

### Potential solution
During PCR, the primers can form dimers to produce a byproduct around 100 bp. As long as the band is not dominant, and the amount of cDNA is good, this will not affect the end-result. However, the band may become dominant, if your mRNA quality is poor, or the concentration is too low. Make sure the quality of you RNA is sufficient and the concentration is 20 ng/μL.

### Problem 3
The concentration of the ready library is low (below 2 nM) (steps 61 and 62).

### Potential solution
Usually 10 cycles in KIN-P2 (step 50) is enough to give sufficient yield for sequencing. However, if needed, one or two additional cycles can be added to the amplification.

### Problem 4
Low yield of the sequencing result (sequence data).

### Potential solution
Illumina's Sequencing Analysis Viewer (https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html ) helps us dissect the problem; also for the troubleshooting, we start the data processing from the BCL files instead of the FASTQ files. For example, the sequencing quality dropped around the polyG part of the TSO, PhiX concentration should be increased.

### Problem 5
Biases in the expression data between the libraries (sequence data).

### Potential solution
The sample layout must be considered carefully when one study uses two or more libraries. For example, to reduce the biases, NBGLM-LBC (Katayama et al., 2019) can be applied. Also, as described in steps 61 and 62, the concentrations of the ready libraries are an important checkpoint before the sequencing.

### Problem 6
Reads in the negative control (sequence data).

### Potential solution
Reads in the negative control point to a contamination in the PCR reaction. Follow general recommendations for PCR setup. Dedicate a separate work area for pre-PCR work and decontaminate the area after each use. We also recommend use of pipette tips with aerosol barriers.

## RESOURCE AVAILABILITY

### Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sini Ezer (sini.ezer@helsinki.fi).

### Materials availability
This study did not generate new unique reagents.

### Data and code availability
The data generated during this study will be available upon request.

## CONSORTIA

## ACKNOWLEDGMENTS

## AUTHOR CONTRIBUTIONS

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. Nat. Protoc. 7, 813–828.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 11, 163–166.

Katayama, S., Skoog, T., Söderhäll, C., Einarsdottir, E., Krjutškov, K., and Kere, J. (2019). Guide for library design and bias correction for large-scale transcriptome studies using highly multiplexed RNAseq methods. BMC Bioinformatics 20, 418.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915.

Krjutškov, K., Koel, M., Roost, A.M., Katayama, S., Einarsdottir, E., Jouhilahti, E.M., Söderhäll, C., Jaakma, Ü., Plaas, M., Vesterlund, L., et al. (2016). Globin mRNA reduction for whole-blood transcriptome sequencing. Sci. Rep. 6, 31584.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Töhönen, V., Katayama, S., Vesterlund, L., Jouhilahti, E.M., Sheikhi, M., Madissoon, E., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Bürglin, T.R., et al. (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. Nat. Commun. 6, 8207.

Wickham, H., and Grolemund, G. (2016). R for Data Science : Import, Tidy, Transform, Visualize, and Model Data (O'Reilly).