Annotation of full-length long noncoding RNAs with capture long-read sequencing (CLS)

Running head: Annotation of full-length lncRNAs using CLS method

Silvia Carbonell Sala (1)*, Barbara Uszczyńska-Ratajczak (1)*, Julien Lagarde (1), Rory Johnson (3,4)[#] & Roderic Guigo (1,2)[#]

1. Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain.
2. Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain.
4. Department of Medical Oncology, Inselspital, University Hospital and University of Bern, 3010 Bern, Switzerland
5. Department of Biomedical Research (DBMR), University of Bern, 3008 Bern, Switzerland

* These authors contributed equally

[#]Co-corresponding authors:
Roderic Guigo roderic.guigo@crg.cat
Rory Johnson rory.johnson@dbmr.unibe.ch

Abstract

Metazoan genomes produce thousands of long-noncoding RNAs (lncRNAs), of which just a small fraction have been well characterized. Understanding their biological functions requires accurate annotations, or maps of the precise location and structure of genes and transcripts in the genome. Current lncRNA annotations are limited by compromises between quality and size, with many gene models being fragmentary or uncatalogued. To overcome this, the GENCODE consortium has developed RNA Capture Long-read Sequencing (CLS), an approach combining targeted RNA capture with third-generation long-read sequencing. CLS provides accurate annotations at high-throughput rates. It eliminates the need for noisy transcriptome assembly from short reads, and requires minimal manual curation. The full-length transcript models produced are of quality comparable to present-day manually-curated annotations. Here we describe a detailed CLS protocol, from probe design through long-read sequencing to creation of final annotations.

Keywords: lncRNAs, next-generation sequencing, NGS, targeted RNA sequencing, CaptureSeq, long-read RNA sequencing, PacBio, Nanopore, genome annotation, GENCODE.

1. Introduction

The vast majority of transcribed sequences in mammalian and vertebrate genomes do not encode proteins and are called noncoding RNAs. Long-noncoding RNAs (lncRNAs) are noncoding RNA transcripts longer than

200 nucleotides [1]. In recent years a growing number of lncRNAs have been functionally associated with various biological and pathological processes, including cancer [2–4]. However, the vast majority (~98%) of them is yet to be functionally characterized. The assignment of lncRNA functions relies on high-quality annotation of their gene structure and boundaries, which is currently lacking.

LncRNAs are particularly challenging to annotate compared to other gene types, due to their characteristics of low expression, low evolutionary conservation, high tissue specificity and unknown sequence-function relationship (contrary to open reading frames in protein-coding genes)[1, 5, 6]. Consequently, present-day lncRNA annotations are far from complete. This incompleteness takes two forms: first, many lncRNA loci are entirely missing from annotations; second, lncRNA annotations are fragmentary and often represent only parts of actual gene structures. In other words, a high proportion of lncRNA annotations lack correct 5' or 3' ends, and their true positions may lie many thousands of base pairs away from the annotated location. The ramifications of this for functional studies are profound [7]. Just one example is the requirement for CRISPR functional screens to design targeting constructs within a window of ~150 bp of the target gene's correct 5' end [8]. Thus, lncRNA gene annotations are a critical resource that presently suffer from serious weaknesses.

RNA-sequencing (RNA-seq), based on next-generation sequencing (NGS) and coupled to computational methods, has driven the rapid growth of lncRNA catalogues [9]. However, in the context of annotation, the conventional form of RNA-seq based on short reads of ~150 bp suffers from severe limitations. Although such methods produce hundreds of millions of reads, lowly-expressed transcripts like lncRNAs are weakly sampled [10]. Moreover, such reads are much shorter than the average lncRNA transcript, and thus must be assembled computationally to build transcript models [11, 12]. Such "transcriptome assembly" is a challenging informatics problem and resulting transcript assemblies suffer from a variety of issues, leading to the incompleteness described above [7, 12]. The problem of transcript incompleteness is reduced in manually curated lncRNA catalogs, e.g. GENCODE or RefSeq [13, 14]. In manual annotation, the lncRNA genes and transcript models are built by human annotators based on non-reconstructed transcriptomic and genomic evidence and according to defined protocols. This precise inspection of lncRNA models produces more confident annotations, which nevertheless suffer from many artifacts present in automatic annotations, e.g. the omission of terminal exons, but at lower rates. Nevertheless, the main weakness of manual annotations is their low throughput compared to automated approaches, and requirement of long-term funding.

The advent of third-generation single molecule long-read sequencing ("TGS") can eliminate the need for transcriptome assembly. By using reads from hundreds to thousands of bases in length, such methods reveal the exon connectivity of individual RNA transcripts [15, 16]. While the problem of assembly is solved, a new challenge is introduced in terms of sensitivity: TGS methods typically have low depth (<10 million reads per flow cell), creating the need to boost the read coverage of genes or transcript of interest [15, 17].

Read coverage can be boosted through the use of targeted capture [18]. The recently described "RNA CaptureSeq" approach employs oligonucleotide capture to enrich a cDNA sample for transcripts of interest, which are then sequenced by standard short-read RNA-seq [19, 20]. Capture is performed by hybridization of cDNA to bead-linked and tiled oligonucleotide probes complementary to target transcript exons. RNA CaptureSeq shows improved sensitivity compared to conventional, unbiased RNA-Seq experiments, and

proves to be effective for discovery of lncRNAs in human and mouse tissues [21, 22]. Interestingly, RNA abundance information is faithfully preserved during the capture process, so that RNA CaptureSeq is capable of sensitive estimates of RNA concentration [21]. Despite these advantages, RNA CaptureSeq relies on conventional short-read sequencing and transcriptome assembly, and suffers from their inherent drawbacks in lncRNA annotation [12, 23].

The limitation of short-read sequencing in RNA CaptureSeq has been recently overcome by replacing it with long-read Pacific Biosciences (PacBio) sequencing, in a method termed Capture LongSeq (CLS) [24, 25]. The use of long-reads allowed to eliminate issues specific to short-read transcript assemblies and substantially improved identification of full-length lncRNA models. CLS improved lncRNA annotations by both detecting new lncRNA loci and improving models for existing ones in human and mouse genomes. In conclusion, this approach delivers lncRNA annotation of manual quality at high throughput.

2. Materials

2.1 Equipment

General molecular biology laboratory equipment is required to execute the CLS protocol. A list of necessary equipment is detailed below:

1. Fume and biological cabins.
2. Water bath.
3. PCR thermal cycler.
4. Set of single / eight channel pipettes and sterile RNase-free pipet tips.
5. Microcentrifuge (with rotor for 1.5-2 ml tubes)
6. QuickSpin minicentrifuge for 0.2 ml and 1.5 ml tubes
7. QuickSpin plates centrifuge
8. Magnetic Bead Separation block, eg DynaMag-96 side (Thermo Fisher Scientific - 12331D)
9. Fragment analyzer, eg Bioanalyzer (Agilent)
10. Micro-volume spectrometer, eg Nanodrop (Thermo Fisher Scientific)
11. Fluorometer for UV absorbance–based quantification, eg Qubit (Thermo Fisher Scientific)
12. SpeedVac concentrator (Savant DNA120 SpeedVac system - Thermo Fisher Scientific)
13. Dry heat block for 1.5 ml tubes (Thermomixer® R – Eppendorf)
14. PIPETMAN Classic set: P1000, P200, P20, P10 and P2 (Gilson)

Equipment required for long-read sequencing: RS II/Sequel from Pacific Biosciences (PacBio) or MinION from Oxford Nanopore Technologies (ONT) third-generation long-read sequencers, or equivalent facility services.

Illumina equipment required for short-read sequencing validation:

1. Covaris instrument (Available models: S220™, M220™, E220™ and LE220™) and related materials for cDNA shearing (microtubes APHA fiber – different formats depending on the Covaris instrument model)
2. Illumina cluster station and sequencer, or equivalent facility Illumina services.

2.2 Reagents and kits

In the CLS protocol, certified gamma-irradiated pure material (nuclease free) disposable plasticware and special Eppendorf polypropylene LoBind tubes are used to avoid material degradation and minimize RNA loss [11]. RNAseZAP^{TM} cleaning reagent (R2020-250ML) is used to remove RNase contamination from bench surfaces, non-disposable plasticware, and laboratory equipment. All buffers and solutions used in this protocol are nuclease-free and molecular grade to avoid sample degradation. The full list of materials, reagents and kits required is detailed below.

2.2.1 Specific kits and reagents (product, Company, reference):

1. RiboZero kit x24 reactions, Epicentre-Illlumina, MRZH11124
2. SMARTer PCR cDNA Synthesis Kit (20 rxns), Clontech , 634926
3. Advantage 2 PCR kit (30 rxns), Clontech, 639207
4. NextFlex barcodes x24 (192 rxn), Illumina, NOVA-514103
5. KAPA LTP Library Preparation kit Illumina (x48 reactions), Roche, 7961898001
6. kapa HiFi Hot Start Ready Mix 100 reactions, Roche, kk2601
7. SeqCap EZ Developer Reagent 1mL (100 reactions), Roche, 6684335001
8. SeqCap EZ Hybridization and Wash Kit (x96 rxns), Roche, 5634253001
9. SeqCap EZ HE-oligo kit A 96 reactions (1-12), Roche, 6777287001
10. SeqCap EZ HE-oligo kit A 96 reactions (13-24), Roche, 6777317001
11. SeqCap EZ Choice XL Library, Roche, 6266363001
12. Agencourt AMPure XP, Beckman Coulter, 082A63881
13. Agencourt RNA Clean XP (40mL), Beckman Coulter, 082A63987
14. m-280 Streptavidin Dynabeads, Invitrogen, 11205D
15. Spike-in RNA variant control mix (SIRV-set 1, Iso Mix: E0, E1, E2), Lexogen (Isogen Life Science), SIRV-set 1
16. ERCC ExFold RNA spike-in mixes, Life Technologies (Thermo Fisher), 4456739

2.2.2 Oligos and blockers (Name, Sequence from 5' to 3', Comments):

1. SMART IIA oligonucleotide, AAGCAGTGGTATCAACGCAGAGTACXXXXX, X = corresponds to undisclosed base in the proprietary SMARTer oligo sequence

2. CDS Primer IIA, AAGCAGTGGTATCAACGCAGAGTAC(T)30 N–1 N, N = A, G, C, or T; N–1 = A, G, or C

3. 5' PCR Primer, AAGCAGTGGTATCAACGCAGAGT

4. TrueSeq Universal adapter, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

5. TrueSeq Indexed adaper, GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-XXXXXX-ATCTCGTATGCCGTCTTCTGCTTG, X = corresponds to Illumina index sequence

6. SMARTer blocker (PolyT blocker), TTTTTTTTTTTTTTTTTTTTTTTTTTTTT
7. SMARTer 5' PCR blocker, AAGCAGTGGTATCAACGCAGAGTAC

8. TrueSeq Universal Blocker, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

9. TrueSeq Indexed Blocker, GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-XXXXXX-ATCTCGTATGCCGTCTTCTGCTTG, X = corresponds to Illumina index sequence

2.2.3 General lab material and reagents:

1. RNAseZAP
2. Ethanol absolute Molecular biology grade
3. Ultrapure DNAsa/RNAsa free distilled water (500mL)
4. RNAse and DNAse free filtered plastic tips
5. RNAse and DNAse free plastic tubs (50 mL, 15 mL, 2 mL, 1.5 mL and 0.2 mL)
6. QubitTM dsDNA BR Assay kit
7. Bioanalyzer chips
8. Lab Gloves

3. Methods

Figures 1 and 2 contain the complete workflow of the CLS protocol.

3.1 General Considerations.

The protocol aims to obtain full-length lncRNAs, making it crucial to maintain RNA integrity by following the general remarks on handling RNA. Wearing latex or vinyl gloves while handling reagents and RNA samples is essential to prevent RNase contamination. Gloves must be changed frequently and the tubes should be closed whenever possible. RNA samples should be kept on ice. All buffers and solutions need to be prepared using nuclease-free water and molecular grade reagents.

The protocol uses harmful chemicals. Pay special attention to always wear a suitable lab coat, disposable gloves, and protective goggles, and follow appropriate regulations when handling chemicals and disposing of waste.

3.2 Design of capture probes

CLS depends on targeting a defined subset of the transcriptome for sequencing – the "target regions". This subset must be defined prior to experimental work, and forms the basis for design of the oligonucleotides comprising the capture library. Thus, the target set might be the entire lncRNA annotation for a given organism, or a subset of it. For example, Lagarde et al. used almost the entire intergenic lncRNA annotation [24], whereas

Deveson et al. targeted an entire chromosome, human Chromosome 21 [25]. Alternatively, suspected candidate transcribed regions might be targeted; for example, Lagarde et al. probed small RNA genes, enhancers and ultraconserved noncoding elements (UCNE) by tiling capture probes across a 1kb window centered on each feature [24]. It is recommended to exclude target regions that overlap, on the same strand, highly-transcribed exons from other genes, such as mRNAs, since these are not of interest but will likely dominate the resulting reads. Similarly, we recommend to remove regions targeting highly-expressed genes, otherwise these will dominate sequencing, and anyway will already be abundantly present as non-specific background in captured cDNA. Identifying such highly-expressed genes introduces the requirement for estimating expression of probed regions in the RNA samples of interest. This may be achieved using in-house or public short-read RNA-seq from the same or similar cell types.

We also recommend adding a number of controls to the design. Depending on the experiment, these could include: protein-coding genes, with steady-state levels matched to the distribution of lncRNAs, as positive controls; random genomic regions of 1 kb from the *E. coli* genome, as negative controls; 100 intergenic regions of 1 kb each with no evidence for any transcriptional or regulatory activity, to estimate background detection; synthetic spike-in RNA of known concentrations, such as ERCC spike-ins or SIRVs (see Note 3: SIRVs), as a means of estimating sensitivity.

After filtering, all targets are combined into a single FASTA file and submitted to the commercial provider (e.g. Roche NimbleGen (Madison, WI)) for probe design and synthesis. In our own work we use the SeqCap EZ Choice XL Library product from NimbleGene. For the CLS experiments the oligonucleotide probes for human and mouse designs directly covered 86.6% and 76.3% of target regions with an estimated 96.1% and 85.0% of target regions successfully targeted, respectively. Summary information for CLS targeted regions in human and mouse is provided in Figure 1B in Lagarde et al. [24].

3.3 Sample preparation

3.3.1 RNA samples.

According to our experience, high quality total RNA should be prepared or purchased from commercial sources to obtain optimal results. The integrity of samples can be tested by evaluating the 28S / 18S rRNA ratio by Bioanalyzer (Agilent): it is recommended to have RNA integrity number (RIN) values of ≥8.5. Sample purity is also checked using ratio of the absorbance at 260 / 280 nm by Nanodrop spectrophotometer (ND-1000 full-spectrum - Thermo Scientific) and quantified with UV absorbance–based quantification Qubit (Thermo Fisher Scientific).

3.3.2 rRNA removal (STEP 1 on workflow figures)

1. Prepare 4 µg of each total RNA sample in a 1.5 ml Eppendorf polypropylene LoBind tube.
2. Add 4 µl of 1:100-diluted Ambion ERCC mix according to the manufacturer's protocol to each total RNA sample (see Note 3: SIRVs).
3. Each total RNA sample containing ERCC controls is ribo-depleted with 10 µl of rRNA removal solution from the Ribo-Zero kit and strictly following manufacturer's protocol.

4. Prepare the Streptavidin beads:
    (a) Aliquot 225 µl of Streptavidin magnetic beads for each sample into a single 1.5 ml tube (i.e. for one capture use 225 µl beads and for "n" samples multiply 255 by the number of samples you want to process).
    (b) Place on the magnet for 1 min.
    (c) Remove and discard SN (supernatant).
    (d) Wash with equal volume of RNAse free water.
    (e) Vortex (medium speed).
    (f) Place on the magnet for 1 min.
    (g) Remove and discard SN.
    (h) Repeat steps from "d" to "g" for a total of 2 washes.
    (i) After removing the the RNAse free water from second wash, resuspend the beads adding 1x the original volume using 1x RNAse free water (i.e. for one capture use 65 µl beads and for "n" samples multiply 65 by the number of samples you are processing).
    (j) Vortex (medium speed).
    (k) Add 1 µl of RiboGuard (RNAse inhibitor).
    (l) Vortex (medium speed).
    (m) Keep at RT (Room Temperature) until required.
5. Prepare the rRNA removal mix to anneal rRNA from the sample to rRNA probes:
    (a) In a tube prepare the sample containing Total RNA and ERCC spike-in Mix for rRNA removal adjusting to a 26 µl final volume.
    (b) Add 4 µl of RiboZero Reaction Buffer and 10 µl RiboZero rRNA removal solution to step 16 tube mixture. Reaction total volume is 40 µl.
    (c) Gently mix by pipetting (10-15 times).
    (d) Incubate the mixture at +68ºC for 10 min.
    (e) Incubate the mixture at RT for 5 min
6. Add 66 µl of Streptavidin washed beads (previously prepared on step 4) to the 40 µl of each sample rRNA removal mix (step 5). Reaction total volume is 106 µl.
7. Gently mix by pipetting (10-15 times).
8. Vortex (medium speed) for 10 sec.
9. Place at RT.
10. Repeat these last 4 steps for each sample.
11. Incubate at RT for 5 min.
12. Vortex (medium speed) for 10 sec.
13. Incubate at 50ºC for 5 min.
14. Quick place on the Magnet and wait 1 min.
15. Carefully collect the SN that contains rRNA free RNA (approx. 85-90 µl) avoiding beads collection (beads contain captured rRNA).

16. Place immediately on ICE.
17. Samples are now in a SAFE STOP POINT. It's possible to store the samples at -80ºC or proceed to purification step.
18. Ribo-depleted RNA is purified using Agencourt RNA Clean XP magnetic beads, purifying with fresh prepared 70% Ethanol and recovering ribo-depleted RNA with 11 µl of RNAse-free water:
    (a) Allow the Agencourt RNA Clean XP magnetic beads to warm to room temperature for at least 30 min before use.
    (b) Vortex magnetic beads before to start for 10 seconds to ensure a homogenous mixture of beads.
    (c) Add 160 µl of beads to each 90 µl of rRNA free RNA from step 16 (corresponding to a ratio of approximately 1.8).
    (d) Gently mix by pipetting (10-15 times).
    (e) Incubate at RT for 15 min.
    (f) Magnet for 1 min.
    (g) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 70 %.
    (h) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 10 µl of RNAse free water.
    (i) Gently mix by pipetting (60 times).
    (j) Incubate at RT for 2 min.
    (k) Magnet for 1 min.
    (l) Carefully collect the SN that contains rRNA free purified RNA (Approx. 10 µl).
19. Successful rRNA removal is validated by Bioanalyzer high resolution fragment analyzer. See an example of expected ribo-depleted RNA Bioanalyzer profile in Figure 3A. The two ribosomal peaks (18S and 28S) will disappear after depletion and the RNA integrity number (RIN), evaluated using the 28S to 18S rRNA ratio, will decrease. The ribo-depleted RNA concentration is low and no quantification is required at this step.
20. Samples are now in a SAFE STOP POINT. It's possible to store the samples at -80ºC or proceed to cDNA synthesis.

3.3.3 cDNA synthesis (STEP 2 & 3 on workflow figures)

The ribo-depleted RNA samples that passed the quality control are used for full-length cDNA synthesis with the SMARTer PCR cDNA synthesis kit following manufacturer's protocol. Two independent cDNA synthesis reactions are carried out for each sample to have sufficient full-length cDNA for subsequent library preparation and capture. This approach allows one to decrease the number of PCR cycles required for second strand synthesis. Ribo-depleted RNA is primed using oligo(dT) and specific SMARTer Stranded RNA-seq technology adaptors for cDNA library construction. The adaptors used in the cDNA library construction sequences (SMART IIA oligonucleotide and CDS primer IIA) are listed in Notes (Note 2).

1. Thaw on ice ribo-depleted RNA samples.

2. Aliquot 3.5 µl of each sample to 2 new PCR tubes.
3. Add 1 µl of 3' SMARTer CDS Primer IIA (12 µM) to each sample. Reaction total volume is 4.5 µl.
4. Gently mix by flicking the tube.
5. Quick spin.
6. Incubate the mix using a thermal cycler:
    (a) Step 1: +73ºC for 3 min
    (b) Step 2: +42ºC for 2 min
7. During step 6 PCR incubation prepare first-strand Master Mix. Assemble the first-strand Master Mix on ice. When assembling the mix for processing multiple samples, prepare an excess volume of approx. 5-10% to allow for complete pipetting. Note that each ribo-depleted sample requires two first strand synthesis reactions. Assemble the mix adding the following components for each reaction sample: 2 µl of 5x First Strand Buffer, 0.25 µl of DTT (100 mM), 1 µl of dNTPs Mix (10 mM), 1 µl of SMARTer IIA Oligo (12 µM), 0.25 µl of RNAse inhibitor, 1 µl of SMART Scribe Reverse Transcriptase (100 U).
8. Once step 6 is finished, quickly proceed to next step adding 5.5 µl of freshly prepared first-strand Master Mix to each sample and carefully mix by pipetting. Reaction total volume is 10 µl. Do not remove the tubes from thermal cycler.
9. Quickly proceed with sample incubation on the thermal cycler:
    (a) Step 3: +42ºC for 90 min.
    (b) Step 4: +70ºC for 10 min.

First-strand RNA obtained from the reaction are used for second-strand synthesis with Advantage 2 PCR kit. Two independent second-strand synthesis reactions are carried out for each first-strand RNA sample to avoid PCR bias. Note that at each original ribo-depleted sample that was split in 2 reactions for first-strand synthesis and reaction has now a total of 4 independent second-strand synthesis reactions. The second-strand synthesis is processed according to the manufacturer's protocol with one important modification: the synthesis cycling protocol specified by the manufacturer has been modified by increasing the extension time from 3 to 6 min to favor the synthesis of long fragments.

10. Assemble the second-strand Master Mix by mixing the following reagents for each first-strand synthesis sample from step 9: 74 µl of RNAse free water, 10 µl of 10x Advantage 2 PCR buffer, 2 µl of 50x dNTPs Mix (10 mM), 2 µl of 5' PCR Primer II A (12 µM), 2 µl of 50x Advantage 2 polymerase Mix. When assembling the mix for processing multiple samples, prepare an excess volume of approx. 5-10% to allow for complete pipetting.
11. Add 90 µl of second-strand Master Mix to 10 µl of first-strand synthesis sample from step 9.
12. Perform the second strand reaction dividing into 2 independent PCR reactions of 50 µl each.
13. Proceed with second Strand PCR cycling conditions:
    (a) Step 1: +95ºC for 1 min.
    (b) Step 2: +95ºC for 15 sec.

(c) Step 3: +65ºC for 30 sec.
(d) Step 4: +68ºC for 6 min.
(e) Step 5: Go to step 2, repeat 17 times (for a total of 18 cycles)
(f) Step 5: +72ºC for 10 min.
(g) Step 6: +4ºC for ∞

14. The two independent second-strand reactions from each sample are pooled and purified using 1:1 ratio ([volume of beads]:[volume of cDNA after PCR reaction]) of Agencourt AMPure XP Beads and performing beads washing steps with freshly prepared 80% Ethanol. Purified cDNA is re-suspended in 22 µl of RNAse free water.
    (a) Allow the Agencourt AMPure XP magnetic beads to warm to room temperature for at least 30 min before use.
    (b) Vortex magnetic beads before to start for 10 seconds to ensure a homogenous mixture of beads.
    (c) Add 100 µl of beads to each 100 µl of second-strand cDNA sample from step 14.
    (d) Gently mix by pipetting (10-15 times).
    (e) Incubate at RT for 15 min.
    (f) Magnet for 1 min.
    (g) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.
    (h) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 22 µl of RNAse free water.
    (i) Gently mix by pipetting (60 times).
    (j) Incubate at RT for 2 min.
    (k) Magnet for 1 min.
    (l) Carefully collect the SN that contains second-strand purified cDNA (Approx. 20 µl).
15. Samples are now in a SAFE STOP POINT. It's possible to store the samples at -20ºC and proceed with an aliquot of 4 µl to 16-18 steps (quality check and quantification).
16. The resulting cDNA is quantified with NanoDrop spectrophotometer. See Figure 3B as an example of expected cDNA NanoDrop quantification and ratios (260/280 and 260/230).
17. The length and quality of library are verified by Bioanalyzer – a high resolution fragment analyzer. See an example of expected cDNA Bioanalyzer profile in Figure 3C.
18. After the quality check on NanoDrop and Bioanalyzer the two independent cDNA synthesis reactions of each sample are pooled and more precisely quantified by Qubit UV absorbance-based fluorometer using Qubit dsDNA BR Assay that specifically quantifies double stranded cDNA.

Before proceeding with the protocol it is strongly recommended to perform one PCR reaction to check for genomic DNA contamination in the sample. We performed PCR reaction to amplify a fragment of human gene SLC14A1 (hg38 chr18:45739177-4573958). The primers were designed to target region containing an intron, which is spliced out in the cDNA. The PCR reaction is expected to give a different amplification

product for cDNA and genomic DNA. In our case the PCR product was 187 bp and 404 bp long for cDNA and genomic DNA, respectively.

3.4 Capture

3.4.1 Pre-Capture Library preparation and PCR amplification (Pre-Capture PCR) (STEP 4 & 5 on workflow figures)

1 µg of cDNA from the preceding step is used for the full-length library preparation (Pre-Capture Library). The enzymes and buffers from Kapa Biosystems low-throughput library prep kit are used for Pre-Capture Library construction.

1. Prepare 1 µg of each sample and adjust to a total final volume of 42 µl in nuclease free water. When working with not fragmented full-length cDNA, the end repair step described on original manufacturer protocol is not necessary.
2. Perform the A-tailing (adenylation step). One adenine is added to blunt cDNA 3′ extremities to improve efficiency of Illumina TruSeq adaptors ligation step.
    (a) Assemble A-Tailing Master Mix: 5 µl of 10x KAPA A-Tailing Buffer, 3 µl of KAPA A-Tailing Enzyme.
    (b) Add 8 µl of A-Tailing Master Mix to each sample. Reaction total volume is 50 µl.
    (c) Gently mix by pipetting (10-15 times).
    (d) Incubate the mixture at 30ºC for 30 minutes.
    (m) During incubation step allow the Agencourt AMPure XP magnetic beads to adjust to the room temperature for at least 30 min before use.
    (e) After incubation step quickly proceed to adenylation cleanup. Vortex magnetic beads before to start for 10 seconds to ensure a homogenous mixture of beads.
    (f) Add 90 µl of beads to each 50 µl of adenylated sample (corresponding to a ratio of approximately 1.8).
    (g) Gently mix by pipetting (10-15 times).
    (h) Incubate at RT for 15 min.
    (i) Magnet for 1 min.
    (j) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.
    (k) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 30 µl of nuclease free water.
    (l) Gently mix by pipetting (60 times).
    (m) Proceed immediately to adapter ligation step keeping the AMPure beads.
3. Ligation of different Illumina TruSeq barcoded adaptor hexamer indexes is used to discriminate each sample. The overall structure of cDNA libraries is represented schematically in Figure 2, STEP 5.
    (a) Assemble ligation Master Mix: 10 µl of 5x KAPA Ligation Buffer, 5 µl of KAPA T4 DNA Ligase.

(b) Add 15 µl of ligation Master Mix to each sample. Reaction total volume is 45 µl.
(c) Add 5 µl of corresponding Adapter-Index (10µM) to each sample. Reaction total volume is 50 µl.
(d) Gently mix by pipetting (10-15 times).
(e) Incubate the mixture at 20ºC for 15 minutes.
(f) During incubation step allow the PEG/NaCl SPRI solution to adjust to the room temperature and keep it protected from the light.
(g) After incubation step quickly proceed to first post-ligation cleanup. Vortex PEG/NaCl SPRI solution before to start for 10 seconds to ensure a homogenous mixture of the buffer.
(h) Add 50 µl of PEG/NaCl SPRI solution to each 50 µl of ligation sample (corresponding to a ratio of 1).
(i) Gently mix by pipetting (10-15 times).
(j) Incubate at RT for 15 min.
(k) Magnet for 1 min.
(l) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.
(m) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 50 µl of nuclease free water.
(n) Gently mix by pipetting (60 times).
(o) Proceed to second post-ligation cleanup. Add 50 µl of PEG/NaCl SPRI solution to each 50 µl of first cleanup sample. Repeat the Adaptor-Index Ligation Cleanup for a second time.
(p) Gently mix by pipetting (10-15 times).
(q) Incubate at RT for 15 min.
(r) Magnet for 1 min.
(s) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.
(t) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 42 µl of nuclease free water.
(u) Gently mix by pipetting (60 times).
(v) Incubate at RT for 2 min.
(w) Magnet for 1 min.
(x) Carefully collect the SN that contains Pre-Capture library (Approx. 40 µl).

4. After adapter ligation, the library is amplified by Linker Mediated PCR (Pre-Capture LM - PCR) for ten PCR cycles under standard low-throughput library preparation Kapa Biosystems PCR conditions, except that the PCR extension step need to be increased to 3 min to allow long fragments to be fully amplified.
   (a) Assemble the Pre-Capture LM – PCR master mix on ice: 50 µl of 2x KAPA HiFi HotStart Ready Mix, 10 µl of Pre LM-PCR Oligos 1 & 2 (5µM)
   (b) Add 60 µl of Pre-Capture LM – PCR master mix to 40 µl of Pre-Capture library from step 5.

(c) Perform the Pre-Capture LM – PCR reaction dividing into 2 independent PCR reactions of 50 µl each.

(d) Proceed with Pre-Capture LM – PCR cycling conditions:

Step 1: +98ºC for 45 sec.

Step 2: +98ºC for 15 sec.

Step 3: +60ºC for 30 sec.

Step 4: +72ºC for 3 min.

Step 5: Go to step 2, repeat 8 times (for a total of 9 cycles)

Step 5: +72ºC for 1 min.

Step 6: +4ºC for ∞

(e) During the PCR enrichment allow the Agencourt AMPure XP magnetic beads to warm to room temperature for at least 30 min before use.

(f) Proceed with Pre-Capture LM-PCR cleanup by pooling the 2 PCR reactions of each sample.

(g) Vortex magnetic beads before to start for 10 seconds to ensure a homogenous mixture of beads.

(h) Add 100 µl of beads to each 100 µl of pooled Pre-Capture LM-PCR products.

(i) Gently mix by pipetting (10-15 times).

(j) Incubate at RT for 15 min.

(k) Magnet for 1 min.

(l) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.

(m) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 32 µl of RNAse free water.

(n) Gently mix by pipetting (60 times).

(o) Incubate at RT for 2 min.

(p) Magnet for 1 min.

(q) Carefully collect the SN that contains enriched Pre-Capture LM-PCR library (Approx. 30 µl).

(r) Samples are now in a SAFE STOP POINT. It's possible to store the samples at -20ºC and proceed with an aliquot of 4 µl to steps 5 and 6 (quality check and quantification).

5. The quality and length of libraries is checked with Bioanalyzer. See an example of expected Pre-Capture library Bioanalyzer profile in Figure 3D.
6. QubitTM dsDNA BR Assay kit is used for precise library quantification.

3.4.2 Pre-Capture sample pooling (Optional)

Pre-capture multiplexing allows one to pool samples prior to capture and is ideal when the number of capture reactions is limited. As recommended by NimbleGen (Roche), the final amount of cDNA in the pooled samples is 1 µg. An equal mass of each sample should be combined in the pool, as estimated by measured concentrations. Salts may affect cDNA capture; therefore, it is crucial to dilute samples with water during the normalization and pooling.

3.4.3 cDNA capture (STEP 6 & 7 on workflow figures)

As indicated by the manufacturer's protocol, denaturation is a key prerequisite for optimal capture performance, therefore it is very important to allow the heat block to reach +95°C and equilibrate prior to capture reaction.

1. Remove the appropriate number of 4.5 µl SeqCap EZ Library aliquots (1 per sample / pool library to capture) from the -15° to -25°C freezer and allow them to thaw on ice.
2. Prepare 1 µg of each sample / pool library on 1.5 ml tube.
3. After normalization, a fixed volume of 4 µl of blocking mix is added to each normalized sample according to the manufacturer's protocol. Assemble the blocking mix with:
    (a) 1 µl of Cot1 DNA (5 µg) for human samples and 1 µl of SeqCap EZ Developer reagent (5 µg), not included in the standard capture kit, for mouse samples. These reagents block repetitive sequences in the cDNA and are species specific.
    (b) 1 µl of TS-HE universal blocking oligo 1,000 pmol. This oligo blocks universal adapters added during ligation step. (see Note 2).
    (c) 1 µl TS-HE adapter-index specific blocking oligo 1,000 pmol. This oligo blocks adapter-index specific added during ligation step. TS-HE adapter-index specific blocking oligo are standard Illumina 6-mer indexes (see Note 2).
    (d) 1 µl SMARTer adapters specific blocking oligonucleotides 1,000 pmol. Please note that these are additional blocking oligos not described in the original manufacturer's protocol. In our previous work we found that addition of these blockers is essential for high on-target enrichment in capture reactions [24] (see Note 2).
4. Dry cDNA hybridization mix from step 3 in a nucleic acid vacuum concentrator on high heat (+60 °C). Denaturation of the cDNA with high heat is not problematic after linker ligation because the hybridization utilizes single-stranded cDNA.
5. Add 7.5 µl 2x Hybridization Buffer and 3 µl of Hybridization Component A (supplied in the NimbleGen SeqCap EZ Hybridization and Wash Kit) and proceed with prepared cocktail to denaturation step (+95°C heat block for 10 minutes to denature the cDNA), strictly following manufacturer's protocol specifications.
6. Quick spin the tube and transfer the cocktail to the 4.5 µl aliquot of SeqCap EZ Library in a 0.2 ml PCR tube and proceed to incubation step in a thermocycler at +47°C for 20 hours. We have tested different incubation times (20 – 72 hours), and found that 20h and yields optimal on-target results [24]. The thermocycler's heated lid should be turned on and set to maintain +57°C (10°C above the hybridization temperature).

3.4.4 Washing and Recovering Captured cDNA (STEP 8 on workflow figures)

We strongly recommend to avoid use of dry heat blocks during washing steps [11]. Instead we recommend water baths, for which the temperature can be reliably monitored and maintained at +47°C throughout all

washing steps. We also recommend using an additional, calibrated thermometer in the water bath. It is required to equilibrate buffers at +47°C for at least 2 hours before washing the captured cDNA. Proper washing is crucial for removing off-target regions.

The washing and recovery of captured cDNA should strictly follow the recommendations in the Roche protocol. We introduced only one modification to the original protocol, after hybridization, human and mouse samples / pools are washed with m-280 Streptavidin Dynabeads to eliminate nonspecific hybridization instead of using Capture Beads provided in the kit. The m-280 Streptavidin Dynabeads have high binding capacity and they are recommended for small biotinylated probe hybridized with long non-biotinylated cDNA [24].

1. Prepare the washing buffers mixing:

1X Washing Buffer preparation	**µl 1x**
10X Stringent Wash Buffer + Nuclease free Water	40 + 360
10X Wash Buffer I + Nuclease free Water	30 + 270
10X Wash Buffer II + Nuclease free Water	20 + 180
10X Wash Buffer III + Nuclease free Water	20 + 180
2.5X Bead Wash Buffer + Nuclease free Water	200 + 300

2. Preheat for each sample the following wash buffers to +47 °C in a water bath: 400 µL of 1X Stringent Wash Buffer and 100 µl of 1X Wash Buffer I.
3. Prepare the Capture Beads:
    (a) Mix m-280 streptavidin beads thoroughly by vortexing for 15 seconds.
    (b) Aliquot 100 µl of m-280 streptavidin beads in a 1.5 ml tube for each sample.
    (c) Magnet for 2 min.
    (d) Remove and discard SN (supernatant).
    (e) Add 200 µl of 1X Bead Wash Buffer.
    (f) Vortex (medium speed).
    (g) Magnet for 2 min.
    (h) Repeat a second wash with 200 µl of 1X Bead Wash Buffer.
    (i) Add 100 µl of 1X Bead Wash Buffer.
    (j) Vortex (medium speed).
    (k) Magnet for 2 min.
    (l) Remove and discard SN and proceed immediately to Capture Washing steps (Do not allow the Capture Beads to dry out. Small amounts of residual Bead Wash Buffer will not interfere with binding of cDNA to the Capture Beads.).
4. Transfer the cDNA hybridization samples prepared in the previous step (coming from incubation step in a thermocycler at +47°C for 20 hours to the Capture Beads).
5. Mix thoroughly by pipetting up and down 10 times

6. Bind the captured sample to the beads by placing the tubes containing the beads and cDNA in a thermocycler set to +47 °C for 45 minutes (heated lid set to +57 °C).
7. Mix the samples by vortexing for 3 seconds at 15 minute intervals to ensure that the beads remain in suspension
8. Start the washing and recovery steps of the captured cDNA:
    (a) Add 100 µl of 1X Wash Buffer I (heated to 47 ºC). Vortex (medium speed) and place on a magnet until SN is clear. And finally remove and discard SN.
    (b) Add 200 µl of 1X Stringent Wash Buffer (heated to 47 ºC). Gently mix by pipetting (10-15 times). Work quickly so that the temperature does not drop much below +47 °C. Incubate at +47 °C for 5 minutes (using water bath). After incubation place the sample on the magnet until SN is clear. Remove and discard SN
    (c) Exactly repeat the previous step performing a second wash with Stringent Wash Buffer.
    (d) Add 200 µl of 1X Wash Buffer I (RT). Vortex (medium speed) for 2 min and place on magnet until SN is clear. Remove and discard SN.
    (e) Add 200 µl of 1X Wash Buffer II (RT). Vortex (medium speed) for 1 min and place on magnet until SN is clear. Remove and discard SN.
    (f) Add 200 µl of 1X Wash Buffer III (RT). Vortex (medium speed) for 30 sec and place on magnet until SN is clear. Remove and discard SN.
    (g) Resuspend the beads with 50 µl of Nuclease Free Water. Sample need to be keep with the beads to proceed to enrichment of selected cDNA targets performing the Post-Capture LM-PCR.

3.4.5 Amplification (Post-Capture ML-PCR), cleanup and quality control of captured cDNA (STEP 9 on workflow figures)

1. Human and mouse washed samples / pools are PCR-amplified with Kapa HotStart ReadyMix 2X in 50 µl reaction volume. Prepare two independent Post-Capture ML-PCR reactions containing half of the washed sample / pool library each, Kapa HotStart Ready Mix and ML-PCR primer mix. Sample split reduces the number of PCR duplicates. 18 PCR cycles are performed, with an increased extension step of 3 min to allow long fragments to be fully amplified.
    (a) Assemble the Post-Capture LM – PCR master mix on ice: 50 µl of 2x KAPA HiFi HotStart Ready Mix, 10 µl of Pre LM-PCR Oligos 1 & 2 (5µM).
    (b) Add 60 µl of Post-Capture LM – PCR master mix to an aliquot of 40 µl of Post-Capture library. Keep the remaining 10 µl in case more enriched product is required or PCR fail.
    (c) Perform the Post-Capture LM – PCR reaction dividing into 2 independent PCR reactions of 50 µl each.
    (d) Proceed with Pre-Capture LM – PCR cycling conditions:

        Step 1: +98ºC for 45 sec.

        Step 2: +98ºC for 15 sec.

Step 3: +60ºC for 30 sec.

Step 4: +72ºC for 3 min.

Step 5: Go to step 2, repeat 17 times (for a total of 18 cycles)

Step 5: +72ºC for 1 min.

Step 6: +4ºC for ∞

(e) During the PCR enrichment allow the Agencourt AMPure XP magnetic beads to warm to room temperature for at least 30 min before use.

(f) Proceed with Post-Capture LM-PCR cleanup by pooling the 2 PCR reactions of each sample.

(g) Vortex magnetic beads before to start for 10 seconds to ensure a homogenous mixture of beads.

(h) Add 100 µl of beads to each 100 µl of pooled Pre-Capture LM-PCR products.

(i) Gently mix by pipetting (10-15 times).

(j) Incubate at RT for 15 min.

(k) Magnet for 1 min.

(l) Wash the beads 2 times with 200 µl of fresh prepared Ethanol 80%.

(m) Air dry the beads at RT (avoid over drying) and then remove the sample from the magnet and resuspend with 32 µl of RNAse free water.

(n) Gently mix by pipetting (60 times).

(o) Incubate at RT for 2 min.

(p) Magnet for 1 min.

(q) Carefully collect the SN that contains enriched Pre-Capture LM-PCR library (Approx. 30 µl).

(r) Samples are now in a SAFE STOP POINT. It's possible to store the samples at -20ºC and proceed with an aliquot of 4 µl to steps 5 and 6 (quality check and quantification).

2. The length of post-capture PacBio and Illumina libraries is verified by Bioanalyzer (Agilent), and quantity is checked by Qubit. See an example of expected Post-Capture library Bioanalyzer profile in Figure 3D.

3.4.6 Post-capture sample pooling (Optional)

Post-capture multiplexing allows one to pool samples prior to sequencing and is ideal in situations where the cost of sequencing is a limitation. After pooling, samples should be quantified again by Qubit dsDNA BR assay to provide optimal calculations for loading samples into the sequencer.

3.5 Sequencing and data analysis (STEP 10 on workflow figures)

The post-Capture libraries (both individual and pooled) can be stored at -20ºC and are ready for standard long-read sequencing library preparation (PacBio / Nanopore). Produced long reads can be directly mapped to the reference genome with a reference annotation provided as a guide to the aligner to improve the splice junction accuracy. For more details regarding the data analysis, please refer to online methods provided in Lagarde et al. [24].

4. Notes

1. Short-read Illumina sequencing (Optional):
    (a) We recommend to use standard Illumina protocol for this purpose. Please note that produced Illumina CLS libraries are unstranded and need to be Covaris-fragmented before the capture (before Pre-Capture Library preparation step). Future CLS protocol improvements include validation of stranded libraries preparation and capture for sequencing using HiSeq IIlumina instruments.
    (b) At this stage an additional Covaris-fragmented Illumina sequencing library for each cDNA sample (1 μg) is prepared according to standard protocols and using following Covaris parameters: Duty cycle 10%, Intensity 5, Cycles per burst 200, Time 30 seconds and sample volume 55 μl. Once the sample is fragmented by Covaris, standard protocol for Pre-Capture Library construction is used to obtain Illumina barcoded Pre-Capture Library.
    (c) Pre-Capture Library is captured following same CLS protocol described at section 3.4 Capture.
    (d) Post-capture Illumina cDNA libraries for splice junction validation and error-rate correction are sequenced by Illumina HiSeq 2500 machine (2 × 125 nt, v4, high-output mode). A large amount of SMARTer adapters, used for full length libraries construction, are present and this could decrease the library diversity. To avoid this and increase libraries' diversity, it is recommended to sequence samples in the presence of at least 10% of PhiX control DNA. Read pairs were demultiplexed with Illumina software.
2. Primer, adapters and blocking sequences used for the CLS protocol are listed in section 2.2.2. All these oligos should be prepared using HPLC purification.
3. In future we intend to introduce the Lexogen's Spike-In RNA Variants (SIRVs) (https://www.lexogen.com/sirvs/sirv-sets/). SIRVS are exogenous synthetic controls for RNA sequencing experiments that can be used as spike-in alone or together with ERCC. Due to their isoform complexity, SIRVs are useful controls for evaluating the sensitivity and specificity of transcript annotation by CLS.
4. CLS protocol allows for cost reduction by pooling samples, which can be performed at two different stages of the protocol: (1) Before capture. This is ideal solution if the number of available capture reactions is limited. Pooling can be done once the amplified Pre-Capture Library is purified (see Pre-Capture Library preparation and amplification section 3.4.1 and 3.4.2 for further details). (2) After capture. This is recommended solution for reducing the cost of sequencing. Pooling can be done before sequencing, once the amplified post-Capture Library is purified (see Post-Capture Library sections 3.4.5 and 3.4.6 for further details).
5. Barcoding samples is an optional step, but is required to multiplex samples and reduce the cost-effectiveness of captures and/or sequencing. Once the samples are barcoded, after Pre-Capture LM – PCR library preparation (section 3.4.1), they can be pooled before (section 3.4.2) or after capture

(section 3.4.6). We note that the use of these relatively short indexes led to the loss of information during later demultiplexing steps. Improving this issue through the use of standard 16-nt PacBio indexes should be a priority in future versions of CLS.

Figure Legends

Figure 1. CLS workflow part 1. Initial steps of CLS protocol from ribosomal RNA depletion (step 1) to the ligation of barcoded adapters (step 4) are covered. RNA is represented by wavy lines, cDNA by straight lines. **Step 1:** Ribosomal RNA depletion using specific probes (18S and 28S rRNAs are shown in aquamarine and orange, respectively). **Step 2:** Two types of oligonucleotides (highlighted in green) are added to the reaction in order to synthetize the first strand of ribo-depleted RNA: (1) oligo dT annealing to 3'end and (2) template switching oligo (TSO) to secure 5' ends. This step is performed according to SMARTer (Clontech) protocol. **Step 3:** Synthesis of second strand results in obtaining double-stranded cDNA. Targeted sequences are shown in violet, while the off-target regions in grey. **Step 4:** The Illumina TruSeq forked barcoded adapters (universal adapter is shown in blue and index adapter is presented in yellow) are ligated to double-stranded cDNA. Ligation of adapters allows to index and enrich for sequences of interest (that optionally can be pooled before the Capture).

Figure 2. CLS workflow part 2. Further steps of CLS protocol from Pre-Capture PCR (step 5) to production of transcript models (step 10). **Step 5:** During this step the Illumina TruSeq forked adapters become double stranded and the polymerase enriches the molecules within few PCR cycles prior to the Capture step. Each construct contains: Illumina indexed adapter (yellow), TSO SMARTer-Clontech adapter (green), cDNA sequence (purple for targets and grey for off-targets), polyA SMARTer-Clontech adapter (green) and finally Illumina universal adapter (blue). At this stage samples can be optionally pooled to decrease the number of capture reactions. **Step 6:** Probes designed to target sequences of interest (purple) are used to enrich for them. **Step 7:** Capture reaction requires incubating samples for 20 hours at 47ºC with the capture probes. **Step 8:** The capture probes contain biotin that allows to pool-down the targeted sequences by streptavidin coated magnetic beads during the Post-Capture washes. **Step 9:** Post-Capture PCR amplifies captured sequences. After PCR, samples can optionally be pooled to decrease the number of sequencing runs. **Step 10:** The long-read data is used to build transcript models. Rectangles with lighter shading and dashed outlines denote novel exons.

Figure 3. Quality controls (QC). The figure shows the expected QC profiles for different steps during CLS sample preparation. **A.** Typical Bioanalyzer RNA sample profile before (total RNA) and after rRNA depletion. Total RNA before rRNA removal step was run on Agilent RNA 6000 Nano Chip (Agilent - 50671511). The ribo-depleted RNA sample was run on Agilent RNA 6000 Pico Chip (Agilent - 50671513). **B.** Expected profile for full-length cDNA on Nanodrop 2000 (Thermo Scientific). **C.** Typical Bioanalyzer full length cDNA sample profile. The sample was diluted to 5-10 ng/µl and then loaded on DNA High Agilent High Sensitivity DNA chip (Agilent - 50674626). **D.** Expected Bioanalyzer profiles for Pre-Capture and Post-Capture libraries. Samples were diluted to 5-10 ng/µl and then loaded on DNA High Agilent High Sensitivity DNA chip (Agilent

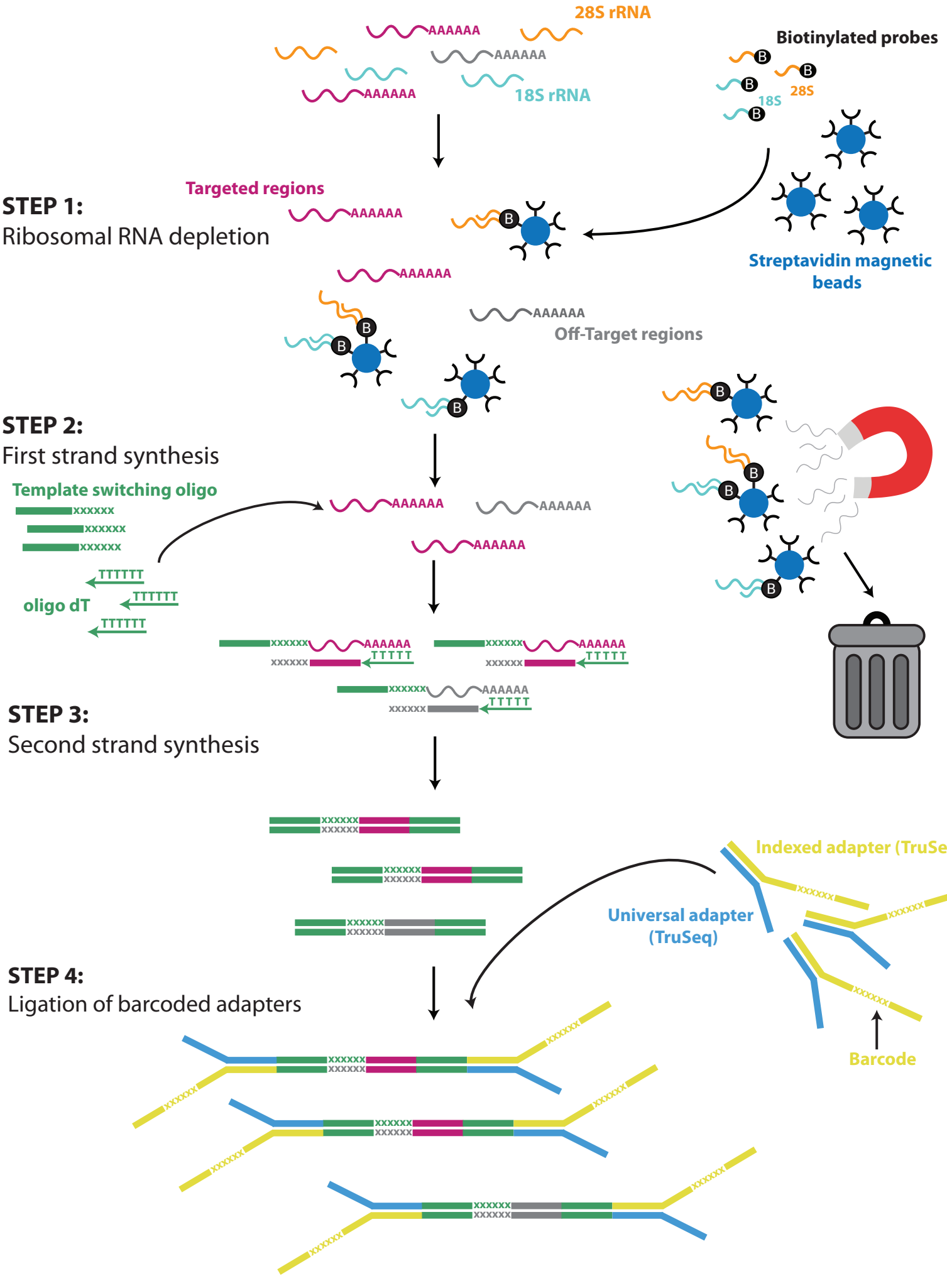
- 50674626). Notice after Pre- and Post-Capture steps (Figure 3D) small cDNA fragments (<300 bp) of full-length cDNA, that can interfere with long read sequencing, observed on Figure 3 C Bioanalyzer are eliminated.

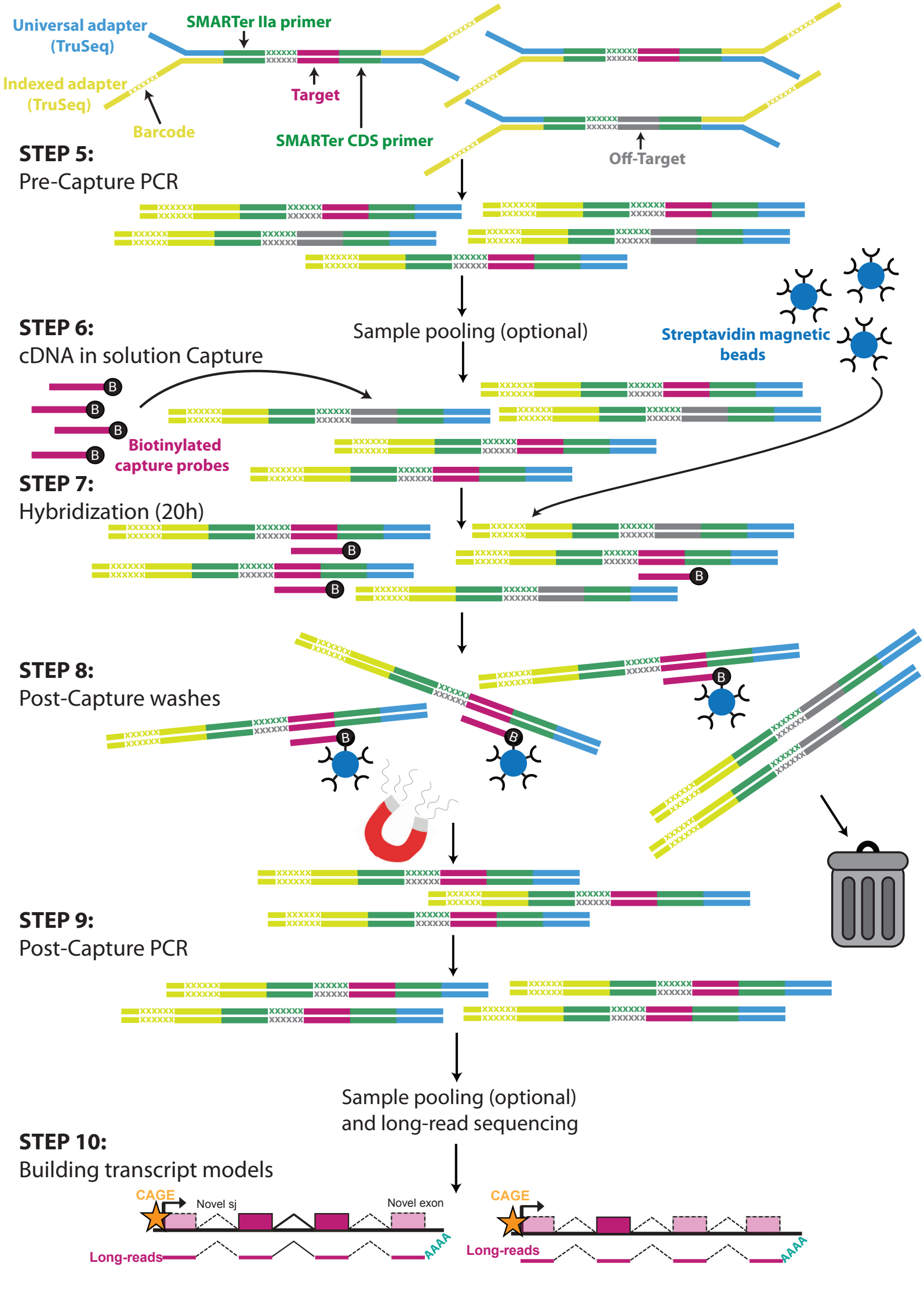
5. References

1. Derrien T, Johnson R, Bussotti G et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–89. doi: 10.1101/gr.132159.111
2. Ulitsky I, Shkumatava A, Jan CH et al (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147:1537–1550. doi: 10.1016/j.cell.2011.11.055
3. Lanzós A, Carlevaro-Fita J, Mularoni L et al (2017) Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. Sci Rep 7:1–16. doi: 10.1038/srep41544
4. Quek XC, Thomson DW, Maag JL V et al (2015) lncRNAdb v2.0: Expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res 43:D168–D173. doi: 10.1093/nar/gku988
5. Mattick JS (2018) The State of Long Non-Coding RNA Biology. Noncoding RNA 4:E17. doi: 10.1007/978-3-319-13689-9_2
6. Hezroni H, Koppstein D, Schwartz MG et al (2015) Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. Cell Rep 11:1110–1122. doi: 10.1016/j.celrep.2015.04.023
7. Uszczynska-Ratajczak B, Lagarde J, Frankish A et al (2018) Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet 19:535–548. doi: 10.1038/s41576-018-0017-y
8. Sanson KR, Hanna RE, Hegde M et al (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun 9:1–15. doi: 10.1038/s41467-018-07901-8
9. Fang S, Zhang L, Guo J et al (2018) NONCODEV5: A comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res 46:D308–D314. doi: 10.1093/nar/gkx1107
10. Lagarde J, Uszczynska-Ratajczak B, Santoyo-Lopez J et al (2016) Extension of human lncRNA transcripts by RACE coupled with long-read high-throughput sequencing (RACE-Seq). Nat Commun. doi: 10.1038/ncomms12339
11. Hansen KD, Brenner SE, Dudoit S (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res 38:1–7. doi: 10.1093/nar/gkq224
12. Steijger T, Abril JF, Engström PG et al (2013) Assessment of transcript reconstruction methods for RNA-seq. Nat Methods 10:1177–84. doi: 10.1038/nmeth.2714
13. Frankish A, Diekhans M, Ferreira A-M et al (2018) GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 1–8. doi: 10.1093/nar/gky955
14. Pruitt KD, Brown GR, Hiatt SM et al (2014) RefSeq: An update on mammalian reference sequences. Nucleic Acids Res 42:756–763. doi: 10.1093/nar/gkt1114
15. Sharon D, Tilgner H, Grubert F, Snyder M (2013) A single-molecule long-read survey of the human

transcriptome. Nat Biotechnol 31:1009–1014. doi: 10.1038/nbt.2705

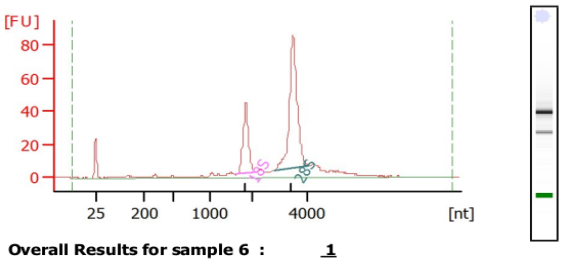
16. Tilgner H, Raha D, Habegger L et al (2013) Accurate Identification and Analysis of Human mRNA Isoforms Using Deep Long Read Sequencing. G3: Genes|Genomes|Genetics 3:387–397. doi: 10.1534/g3.112.004812
17. Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:1–11. doi: 10.1186/s13059-016-1122-x
18. Shyra J. Crider-Miller, Laura H. Reid, Michael J. Higgins, Norma J. Nowak, Thomas B. Shows PAF and BEW (1997) Novel Transcribed Sequences within the BWS/WT2 Region in 11p15.5: Tissue-Specific Expression Correlates with Cancer Type. Genomics 46:355–363. doi: 10.1006/geno.1997.5061
19. Mercer TR, Gerhardt DJ, Dinger ME et al (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol 30:99–104. doi: 10.1038/nbt.2024
20. Mercer TR, Clark MB, Crawford J et al (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc 9:989–1009. doi: 10.1038/nprot.2014.058
21. Clark MB, Mercer TR, Bussotti G et al (2015) Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. Nat Methods 12:339–42. doi: 10.1038/nmeth.3321
22. Bussotti G, Leonardi T, Clark MB et al (2016) Improved definition of the mouse transcriptome via targeted RNA sequencing. Genome Res 26:705–716. doi: 10.1101/gr.199760.115
23. Hardwick SA, Chen WY, Wong T et al (2016) Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13:792–798. doi: 10.1038/nmeth.3958
24. Lagarde J, Uszczynska-Ratajczak B, Carbonell S et al (2017) High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat Genet 49:1731–1740. doi: 10.1038/ng.3988
25. Deveson IW, Brunck ME, Blackburn J et al (2018) Universal Alternative Splicing of Noncoding Exons. Cell Syst 6:245–255.e5. doi: 10.1016/j.cels.2017.12.005


28S rRNA
AAAAAA
AAAAAA
AAAAAA
18S rRNA
Biotinylated probes
28S
18S
Targeted regions
AAAAAA
STEP 1:
Ribosomal RNA depletion
Streptavidin magnetic beads
AAAAAA
AAAAAA
Off-Target regions
STEP 2:
First strand synthesis
Template switching oligo
xxxxxx
xxxxxx
xxxxxx
TTTTTT
oligo dT
TTTTTT
TTTTTT
AAAAAA
AAAAAA
AAAAAA
xxxxxx
AAAAAA
TTTTT
xxxxxx
STEP 3:
Second strand synthesis
Indexed adapter (TruSeq)
Universal adapter
(TruSeq)
STEP 4:
Ligation of barcoded adapters
Barcode


Universal adapter (TruSeq)
SMARTer IIa primer
Indexed adapter (TruSeq)
Target
Barcode
SMARTer CDS primer
Off-Target
STEP 5:
Pre-Capture PCR
Sample pooling (optional)
STEP 6:
cDNA in solution Capture
Streptavidin magnetic beads
Biotinylated capture probes
STEP 7:
Hybridization (20h)
STEP 8:
Post-Capture washes
STEP 9:
Post-Capture PCR
Sample pooling (optional) and long-read sequencing
STEP 10:
Building transcript models
CAGE
Novel sj
Novel exon
Long-reads
AAAA
CAGE
Long-reads
AAAA

A.

Before rRNA removal



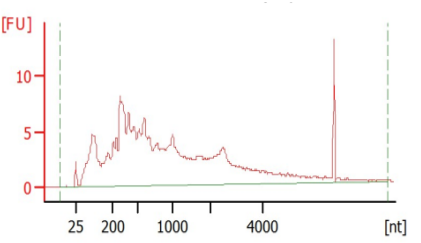
Overall Results for sample 6 : 1

RNA Area:	463.1
RNA Concentration:	148 ng/µl
rRNA Ratio [28s / 18s]:	2.5
RNA Integrity Number (RIN):	9.8 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.80

Fragment table for sample 6 : 1

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,604	2,124	80.3	17.3
28S	2,783	4,047	200.9	43.4

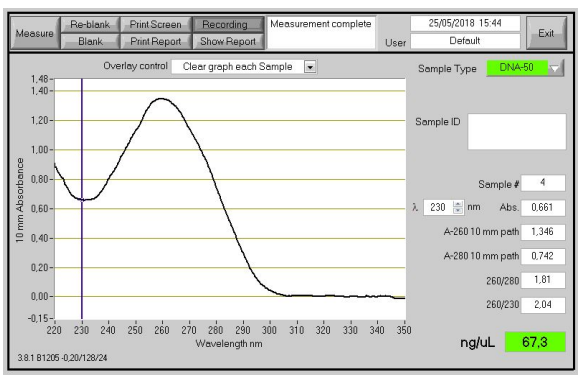
After rRNA removal



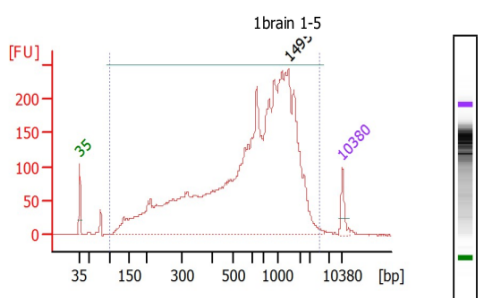
Overall Results for sample 3 : HS Brain 2

RNA Area:	270,8
RNA Concentration:	920 pg/µl
rRNA Ratio [28s / 18s]:	0,0
RNA Integrity Number (RIN):	2.5 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 2.50

B.



C.



Peak table for sample 6 : 1brain 1-5

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	1,495	4,128.95	4,185.3	
3	10,380	75.00	10.9	Upper Marker

Overall Results for sample 6 : 1brain 1-5

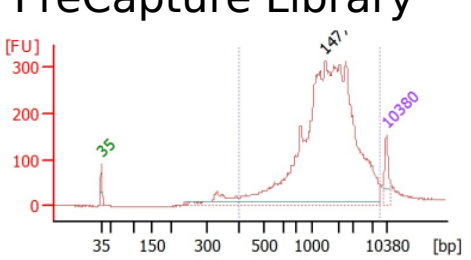
Number of peaks found:	1
Noise:	0.6
Corr. Area 1:	5,816.8

Region table for sample 6 : 1brain 1-5

From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]
100	5,000	881	5,931.19	20,936.3

D.

PreCapture Library



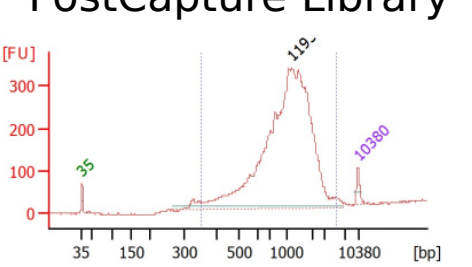
Overall Results for sample 3 : BeforeCAP_smarter

Number of peaks found:	1
Noise:	0.8
Corr. Area 1:	4,449.9

Region table for sample 3 : BeforeCAP_smarter

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
397	8,504	4,449.9	94	2,012	74.2	2,483.43	3,042.0	

PostCapture Library



Overall Results for sample 4 : AfterCAP_smarter

Number of peaks found:	1
Noise:	0.9
Corr. Area 1:	4,370.6

Region table for sample 4 : AfterCAP_smarter

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
353	5,120	4,370.6	98	1,222	51.1	5,359.26	8,987.2	