Introducing synthetic thermostable RNase inhibitors to single-cell RNA-seq

J.C. Noble[1], Antonio Lentini[1], Michael Hageman-Jensen[2], Rickard Sandberg[2], and Björn Reinius[1, #]

[1]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.

[2]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden.

 # Correspondence: bjorn.reinius@ki.se

Abstract

Single-cell RNA-sequencing (scRNAseq) is revolutionizing biomedicine, propelled by advances in methodology, ease of use, and cost reduction of library preparation. Over the past decade, there have been remarkable technical improvements in most aspects of single-cell transcriptomics. Yet, there has been little to no progress in advancing RNase inhibition despite maintained RNA integrity being critical during cell collection, storage, and cDNA library generation. Here, we demonstrate that a synthetic thermostable RNase inhibitor yields single-cell libraries of equal or superior quality compared to ubiquitously used protein-based recombinant RNase inhibitors (RRIs). Importantly, the synthetic RNase inhibitor provides additional unique improvements in reproducibility and throughput, enables new experimental workflows including retained RNase inhibition throughout heat cycles, and can reduce the need for dry-ice transports.

In summary, replacing RRIs represents a substantial advancement in the field of single-cell transcriptomics.

Introduction

The capture of intact RNA is required for RNAseq methods to accurately record the transcriptome of the analyzed sample material. scRNAseq is particularly sensitive to RNA degradation and losses due to the miniscule copy numbers of individual transcripts present in the cell[1,2]. Thus, the inclusion of RRIs, which are *in vitro* synthesized RNase-binding proteins, is nearly universal during cell capture, storage, cell lysis, and reverse transcription (RT) in scRNAseq protocols. However, the use of RRIs is inconvenient not only due to its relatively high cost fraction of the scRNAseq libraries but also due to RRI degradability, which can introduce batch variation in library yield and quality with production lot, storage time, and temperature exposures of the inhibitor. Furthermore, thermal sensitivity of RRIs imposes restrictions on the temperature ranges permissible in experimental protocols and laboratory procedures, while also complicating the logistics of exchanging reagents and biological samples. Due to their unmatched performance, RRIs continue to be the leading standard in scRNAseq, and replacing them has so far proven non-trivial. In certain scRNAseq protocols, developers have substituted RRIs with alternatives like guanidinium salts, beta-mercaptoethanol, or diethyl pyrocarbonate[3,4]. While these chemicals decrease RNA degradation, their performance in scRNAseq is not on par with RRIs. They are furthermore toxic and carry characteristics that are counterproductive in subsequent steps of scRNAseq library preparation, including their chaotropic nature, reduction capability, or susceptibility to temperature and buffer variations. These properties adversely impact RT and PCR processes, ultimately compromising the quality of the sequencing library. To offer a full substitute for RRIs without sacrificing library quality, we developed SEQURNA, a synthetic thermostable RNase inhibitor which consists of a proprietary mix of non-toxic molecules that furthermore do not depend on any toxic reducing agents that are essential for conventional RRI functionality (e.g. DTT or beta-mercaptoethanol).

Here, we demonstrate that RRIs can effectively be replaced by a synthetic thermostable RNase inhibitor across various scRNAseq applications, yielding retained or improved sequencing library characteristics. This enables various novel workflows and logistical benefits impacting single-cell genomics.

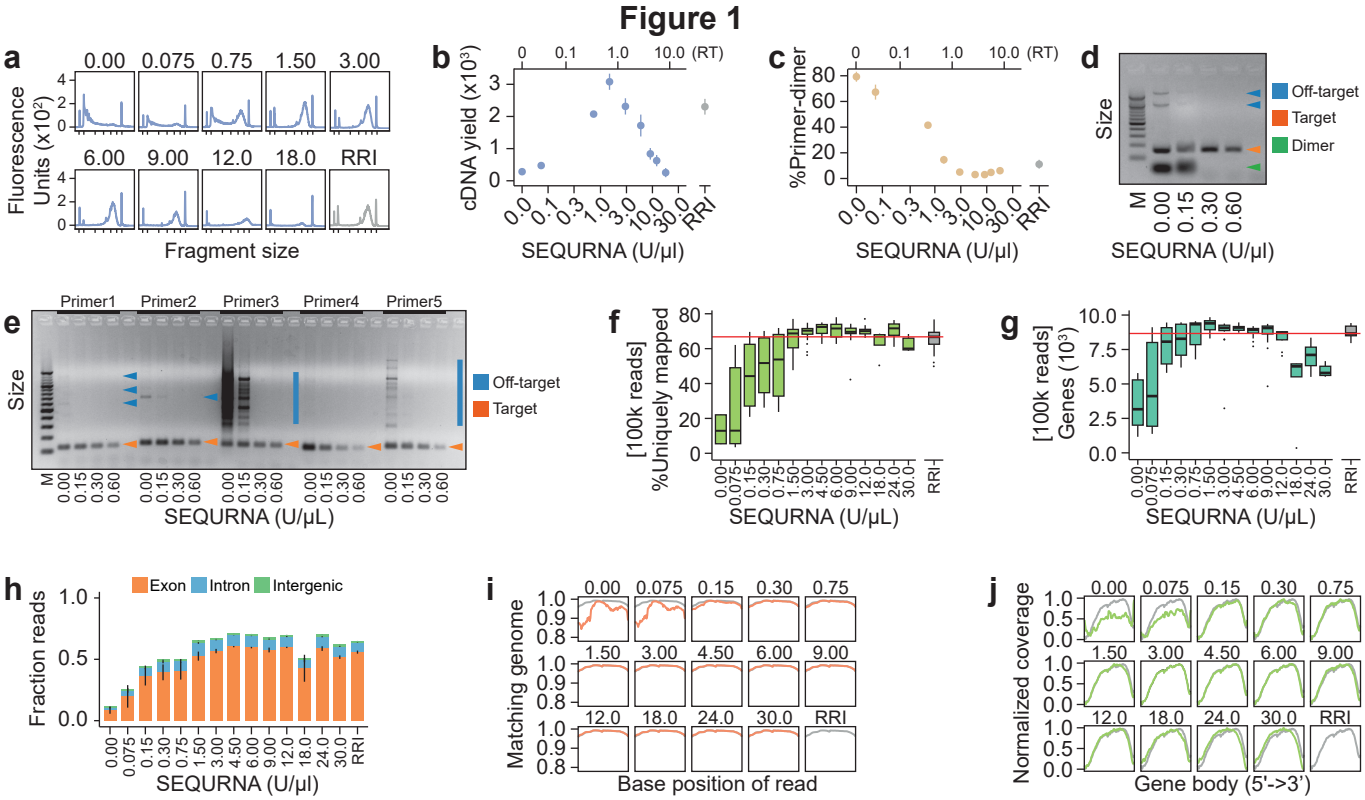


Figure 1. Replacing recombinant RNase inhibitor with a synthetic thermostable RNase inhibitor in mini-bulk RNA-seq library generation.

a. Bioanalyzer traces of Smart-seq2 (SS2) cDNA libraries generated from 100 pg mouse RNA using varying SEQURNA concentrations (0–18 U/µl) or standard RRI in lysis buffer. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **b.** Quantified cDNA yield of bulk SS2 cDNA libraries (fragment size 200–10,000 bp). SEQURNA concentration in lysis buffer on the lower x-axis and resulting concentration in RT on the upper x-axis. Data shown as mean (dot) and standard error (line). n = 2-8 traces per condition (54 traces analyzed). **c.** Percent primer-dimer in SS2 bulk cDNA libraries (fragment size 20–50 bp) for the samples in (**b**). **d.** Gel electrophoresis image of products formed using an eGFP plasmid as template and varying the concentration of SEQURNA during PCR. **e.** Gel electrophoresis image of products formed using human gDNA as template and five primer sets targeting *TUBA1A*. **f-g.** Box plots of percent uniquely mapped sequencing reads to the reference genome (**f**) and number of genes detected (**g**) for SS2 libraries generated from 100 pg mouse RNA. Data shown as median, interquartile range (IQR) and 1.5x IQR for n = 5-35 replicates per condition (total= 195). **h.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the mouse genome for bulk SS2 libraries, with lines indicating standard error. **i.** Line plot of fraction of bases along sequencing reads matching the reference transcripts, indicating read quality, for bulk SS2 libraries. **j.** Line plot of normalized coverage along transcripts for bulk SS2 libraries.

Results

To assess its utility in scRNAseq library preparation, we started by systematically testing concentrations ranges of SEQURNA in the Smart-seq2 (SS2) protocol[5], which provides full-length read coverage across transcripts enabling evaluation of polymerase processivity. The standard SS2 protocol includes a cell lysis and RNA denaturation step which involves heating the sample to 72°C before RT, requiring the addition of fresh RRI in the RT mix due to thermosensitivity of the protein-based RNase inhibitor. Conversely, the synthetic RNase inhibitor was added only in the lysis buffer, retaining effectiveness throughout the denaturation step and eliminating the need for its addition in the RT mix. We experimented with lysis buffers containing a range of SEQURNA (0–30 U/µl; resulting in 0–13.5 U/µl in RT) and 100 pg spiked mouse RNA, performed SS2 library amplification, and evaluated the resulting cDNA by capillary electrophoresis. Within a defined concentration range of SEQURNA (1.5–6 U/µl) the cDNA fragment size distribution was on par with the standard SS2 protocol with RRIs, whereas the cDNA yield was similar or increased with SEQURNA (**Fig. 1a–b**). At lower and higher SEQURNA concentrations, the cDNA yield declined due to RNA degradation and RT inhibition, respectively. Intriguingly, we noticed that SEQURNA also increased PCR specificity in SS2, reducing the fraction of primer-dimer content in the cDNA libraries (**Fig. 1a, c**). We further explored this novel property of SEQURNA by a previously established dimer assay using primers intentionally designed to self-hybridize[6] and an eGFP plasmid as template (**Methods**). At increasing SEQURNA concentration, primer-dimers were indeed eliminated whereas the target amplicon was preserved (**Fig. 1d**). We also designed *TUBA1A*-targeting primer pairs of varying tendencies to form unspecific products and found that SEQURNA could prevent unintended bands and smear products amplified from human gDNA (**Fig. 1e**). By melting-curve analysis, we found that the increased PCR stringency was not the mere result of altering melting temperature (**Supplementary Fig. 1a**). Notably, enhanced stringency may increase the proportion of informative fragments in scRNAseq libraries (**Fig. 1c**). To examine the information yield and data quality of the SS2 libraries, we sequenced libraries generated from mouse RNA in presence of SEQURNA or standard RRI, and downsampled the resulting reads to allow even comparison (**Methods**). We analyzed library quality metrics in terms of genomic and transcriptomic mappability, base substitution rates, insertion and deletions (indels), likelihood of mismatches within the read, and average indel length. SEQURNA conditions 1.5–12 U/µl in the lysis buffer (0.675–5.4 U/µl in RT) yielded libraries of similar quality as RRI with respect to all parameters investigated, including the number of detected genes, fraction reads mapped to exons, base-along-read likelihood of matching genome, and gene body coverage (**Fig. 1f–j** and **Supplementary Fig. 1b, c**). Additionally, substitution and indel rates were on par between SEQURNA and RRI in the relevant concentration range (**Supplementary Fig. 1d–i**), consistent with unaffected polymerase fidelity.

Together, these results demonstrate a successful working range (~1 order of magnitude) of SEQURNA in SS2 library generation and that the cDNA yield curve correlated with downstream quality metrics. Therefore, the notion of a defined working range of the synthetic RNase inhibitor is a key consideration, to be identified for individual scRNAseq protocols as they differ in buffer composition, volumetric changes between cell lysis and RT reaction, and reverse transcriptase used.

A synthetic and thermostable RNase inhibitor might better tolerate variable storage conditions, thus we explored the efficacy of SEQURNA in SS2 library generation after first subjecting inhibitor stocks to various "harsh" treatments (**Methods**) prior to use in lysis buffers, such as heating it to 37°C or 50°C for 24 hours, freeze-thawing up to six times, vortexing for 24 hours, and drastically altering pH (pH 4-10), remarkably observing intact cDNA traces in each treatment (**Supplementary Fig. 2**). Importantly, the robustness of the synthetic inhibitor enables a wide array of experimental treatments and conditions, expanding the range of potential applications and introducing multiple new experimental pathways in single-cell omics techniques.

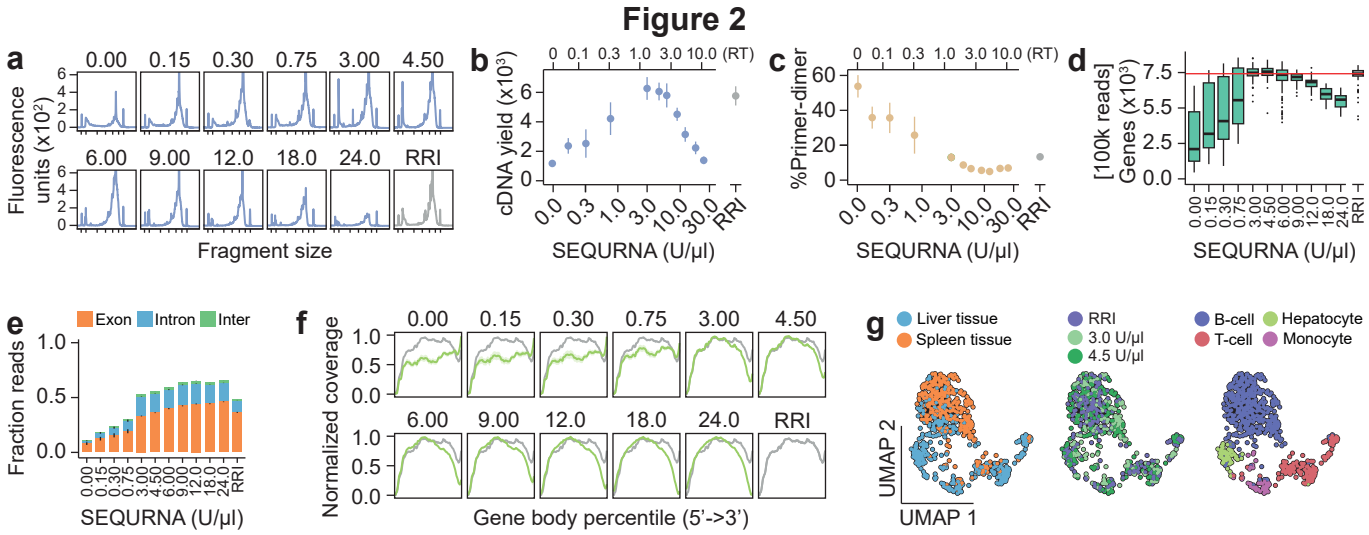


Figure 2. Performance of SEQURNA in single-cell Smart-seq2.
a. Bioanalyzer traces of Smart-seq2 (SS2) cDNA libraries from HEK293FT single cells using varying SEQURNA concentrations (0–24 U/µl) or standard RRI in lysis buffer. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **b.** Quantified cDNA yield of single-cell SS2 cDNA libraries from HEK293FT cells (fragment size 200–10,000 bp). SEQURNA concentration in lysis buffer on the lower x-axis and resulting concentration in RT on the upper x-axis. Data shown as mean (dot) and standard error (line). n = 3-6 traces per condition (50 traces analyzed). **c.** Percentage primer-dimer of single-cell SS2 cDNA libraries from HEK293FT cells (fragment size 20–50 bp) for the samples in (**b**). **d.** Box plot of number of genes detected in HEK293FT single-cell SS2 libraries. Data shown as median, interquartile range (IQR), and 1.5x IQR. n = 37-94 per condition (704 libraries analyzed). **e.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the human genome for single-cell HEK293FT SS2 libraries, with lines indicating standard error, n = 37-94 per condition. **f.** Line plot of normalized coverage along transcripts for single-cell HEK293FT SS2 libraries, n = 37-94 per condition. **g.** Uniform manifold approximation and projection (UMAP) for single-cell SS2 libraries in mouse cells derived from liver (n = 349 cells) and spleen (n = 368 cells) tissue. Cell sorting was gated for small cell size to enrich cells with low RNA content (**Supplementary Fig. 5**).

Encouraged by the results on purified RNA, we FACS sorted individual cultured human embryonic kidney cells (HEK293FT) into 96-well plates containing lysis buffer and 0–24 U/µl SEQURNA, or standard SS2 buffer with RRI (**Methods**), and analyzed the resulting single-cell SS2 libraries. At optimal SEQURNA concentration (peaking around 2–3 U/µl in lysis buffer) (**Fig. 2a–c**), the sequencing libraries and generated transcriptome data were on par with standard SS2 libraries generated using RRI (**Fig. 2d–f** and **Supplementary Fig. 3**), with similar capture of biological signals, exemplified by the cell-cycle (**Supplementary Fig. 4**). Together, these results demonstrate successful capture of minute biological signals in scRNAseq data generated with a synthetic RNase inhibitor, on par with RRIs.

We then challenged the synthetic RNase inhibitor by isolating small cells (low RNA content) dissociated from mouse spleen and liver (**Supplementary Fig. 5a, b**) – tissues known to be of high and moderate RNase content, respectively. Single cells were sorted into lysis buffer containing either RRI or SEQURNA (3 or 4.5 U/µl). The resulting library quality of the three conditions were highly similar (**Supplementary Fig. 5c–p**). We visualized the relationship between cells and tissue origin by Uniform Manifold Approximation and Projection (UMAP) and observed that the distribution of SEQURNA- and RRI-treated samples was well represented throughout the projection and indicating little or no difference in gene and cell-type detection between the conditions (**Fig. 2g**). In brief, independent of inhibitor type, we identified four distinct clusters representing B-cells, T-cells, monocytes/macrophages (clusters 0,1,3) as well as hepatocytes (cluster 2) (**Supplementary Fig. 6**), in line with liver and spleen lymphocyte/monocyte populations identified in previous studies[7,8], demonstrating successful transcriptome capture of small cells from RNase-rich tissue.

To demonstrate broad utility of the synthetic RNase inhibitor in next-generation sequencing protocols, we applied SEQURNA to Smart-seq3 (SS3)[9] on purified mouse RNA and single HEK293FT cells. At identified optimal concentration (0.06-0.6 U/µl SEQURNA in the lysis buffer; 0.03-0.3 U/µl during RT) it produced maximum cDNA yield (**Fig. 3a–d**) and sequencing data of parallel or superior quality metrics as standard SS3 using RRI (**Fig. 3e–g** and **Supplementary Fig. 7**), including similar UMI counts (**Fig. 3h**). Using SEQURNA in one commercial bulk RNAseq kit (TruSeq RNA Library Preparation Kit v2) and two commercial single-cell RNAseq kits (NEBNext Single Cell/Low Input RNA Library Prep Kit, and QIAseq FX Single Cell RNA Library Kit) demonstrated SEQURNA libraries of high quality (**Supplementary Fig. 8**). This implicates broad applicability of the synthetic RNase inhibitor across diverse experimental conditions and RNAseq methods.

A thermostable RNase inhibitor might offer unique RNA integrity protection during sample handling, from the collection of cells until stable cDNA is produced. Based on our results, we hypothesized that cells collected in SEQURNA lysis buffer would be more tolerant to extended waiting times compared to conventional RRIs, a feature of importance for large-scale scRNAseq projects. To this end, cells sorted into SS2 lysis buffer and 0-12 U/µl SEQURNA were stored at ambient temperature (25°C), cold (4°C) or ultra-low temperature (-80°C) for 1, 4, 7, or 14 days before SS2 libraries were generated. Expectedly, we observed that cells stored in lysis buffer lacking inhibitor resulted in degraded cDNA traces (**Supplementary Fig. 9a–d**). Intriguingly, cDNA profiles obtained from SEQURNA samples even at 4 days of storage at room temperature (25°C) showed lower amounts of degradation (**Supplementary Fig. 9c**) with workable integrity even up to day 4, and remarkably, throughout the full 14-day experiment when stored refrigerated (4°C) (**Supplementary Fig. 9d**). Although this observation based on cDNA traces alone should not be interpreted as a complete lack of RNA degradation over storage time, it does indicate a surprising stability of cellular RNA over extended time when kept in lysis buffer containing synthetic inhibitor.

To systematically characterize the capacity of SEQURNA to extend quality in longer-term stored RNA from single cells, we conducted a large-scale

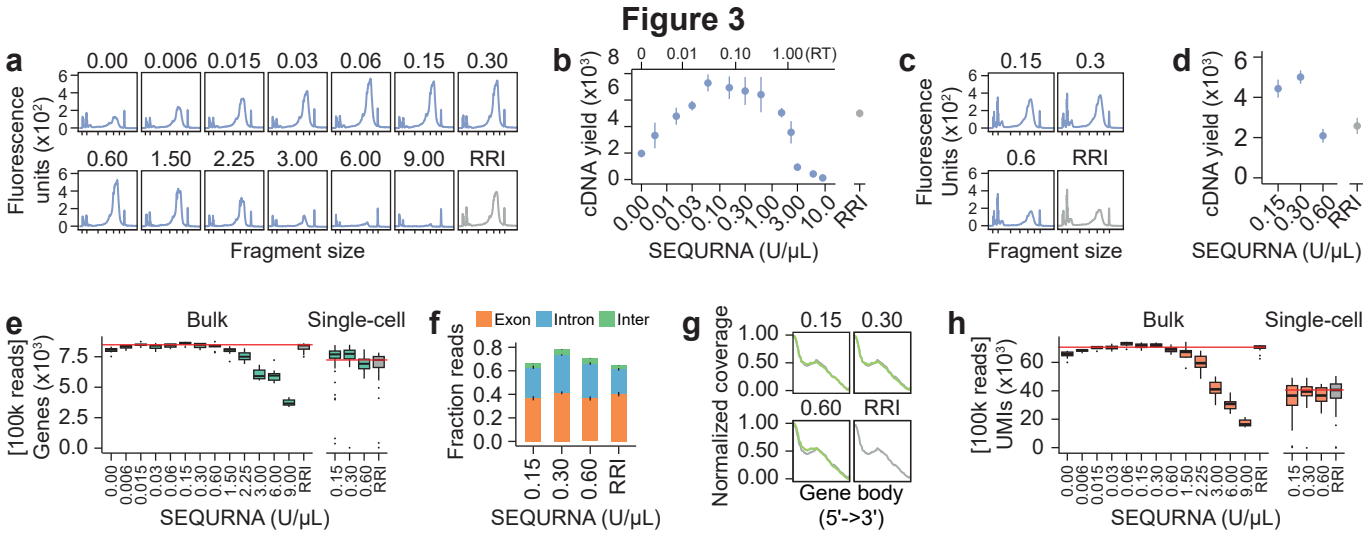


Figure 3. Performance of SEQURNA in mini-bulk and single-cell Smart-seq3.
a. Bioanalyzer traces of Smart-seq3 (SS3) cDNA libraries generated from 100 pg mouse RNA (bulk) using varying SEQURNA concentration (0–9 U/µl) or standard SS3 RRI in lysis buffer. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **b.** Quantified cDNA yield of bulk SS3 cDNA libraries (fragment size 200–10,000 bp). Data shown as mean (dot) and standard error (line). n = 3-5 traces per condition (51 traces analyzed). **c.** Bioanalyzer traces of single-cell SS3 cDNA libraries from HEK293FT cells using varying SEQURNA concentrations (0.15, 0.3, 0.6 U/µL) or standard SS3 RRI in lysis buffer. **d.** Quantified cDNA yield of single-cell SS3 cDNA libraries from HEK293FT cells (fragment size 200–10,000 bp). Data shown as mean (dot) and standard error (line). n = 11-14 traces per condition (53 traces analyzed). **e.** Box plot of number of genes detected in SS3 libraries for 100 pg mouse bulk RNA (left, n =7-16 libraries per condition, total= 170) and single-cell HEK293FT (right, n = 31-45 cells per condition, total= 158). Data shown as median, interquartile range (IQR), and 1.5x IQR. **f.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the human genome for single-cell HEK293FT SS3 libraries, with lines indicating standard error. n = 31-45 cells per condition. **g.** Line plot of normalized coverage along transcripts for single-cell HEK293FT SS3 libraries, with characteristic bias towards the 5'-UMI-containing end. n =31-45 cells per condition. **h.** Box plot of number of unique molecular identifiers (UMIs) detected in SS3 libraries for 100 pg mouse bulk RNA (left, n = 7-16 samples per condition) and single-cell HEK293FT (right, n =31-45 cells per condition).

sequencing experiment utilizing the newly developed Smart-seq3xpress protocol (SS3xpress)[10]. This protocol is a miniaturized adaptation of SS3 which not only utilizes a markedly lowered reaction volume that drastically reduces cost but also streamlines the library-preparation process by completely circumventing the cDNA purification step before tagmentation, instead proceeding with tagmentation and indexing PCR immediately after a cDNA dilution step. This necessitates compatibility between transposase Tn5 and SEQURNA in the relevant concentration range. Thus, we experimentally characterized how SEQURNA affected Tn5 tagmentation of double-stranded cDNA. Surprisingly, we observed that SEQURNA was not only compatible with Tn5 activity but that it increased product yield when SEQURNA was added in low concentration (0.0025–0.1 U/µl) (**Fig 4a–b**) and inhibition was only observed above 0.25 U/µl.

Next, we sorted HEK293FT cells into 384-well plates with SS3xpress lysis buffer containing SEQURNA (0, 0.06, 0.15, 0.3, 0.6, 1.5, or 3.0 U/µl) or standard SS3xpress lysis buffer with RRI. To avoid potential across-plate batch effects, all inhibitor conditions were represented within each plate. Two plates were kept at -80°C immediately after cell sorting (representing the day 0 time point) while the other plates were stored at either room temperature (25°C) or refrigerated (4°C) for up to 14 days, and thereafter processed into SS3xpress libraries and sequenced (**Methods**). In the absence of cDNA traces (due to direct tagmentation in SS3xpress), we identified the optimal concentration range of SEQURNA directly from read mapping statistics from day 0, where the optimal range was 0.06–0.3 U/µl SEQURNA in the lysis buffer (resulting in 0.045-0.225 U/µl in RT reaction) (**Fig. 4c–e**), i.e. similar to that identified for SS3. Interestingly, we observed that SEQURNA in the optimal range (0.06–0.3 U/µl) consistently produced improved libraries (more genes and UMIs detected per cell) compared to standard SS3xpress using recombinant RRI, indicating that SEQURNA can enhance the established SS3xpress protocol. Next, we analyzed downsampled data from cells stored at 25°C and 4°C, only considering the working range of SEQURNA concentrations (0.06–0.6 U/µl), no inhibitor (0 U/µl), and standard RRI. As expected, we observed the overall trend of decreasing number of genes and UMIs detected per cell over time in storage, and a more rapid degradation at 25°C compared to 4°C (**Fig. 4f–i**). However, degradation progressed markedly slower for cells in SEQURNA lysis buffer, an effect which was more prominent in the 4°C condition, indicating that the interaction of cooling and SEQURNA was especially beneficial for delaying cell RNA degradation, in line with the previous observations of SS2 cDNA traces.

Thus, we conclude that substituting RRI with the synthetic thermostable RNase inhibitor in the lysis buffer extends the workable time in between collecting and processing cells into scRNAseq libraries. This has important logistical benefits when handling large sample batches, that ultimately may improve information yield in single-cell research.

Discussion

Here, we report that a synthetic RNase inhibitor (SEQURNA) can fully replace gold standard RRIs in multiple scRNAseq and bulk RNAseq applications without compromising library quality. This is likely to transform scRNAseq protocols going forward. As part of this, we provide updated versions of the Smart-seq2, Smart-seq3, and Smart-seq3 express protocols (**Supplementary Notes 1–3**).

The synthetic RNase inhibitor has several advantages. Thermostability and robustness to various other harsh chemical and physical treatments enable more environmentally sustainable transportation and storage of the inhibitor at room temperature as well as novel and simplified workflows that are currently precluded by the less inert protein-based RRIs, which are thermosensitive and generally transported on dry ice. For example, stable lysis buffers and spotted plates can be pre-made, frozen, thawed, or kept at room temperature for extended periods of time and need not be freshly prepared before cell collection, simplifying workflows and across-lab collaborations. These properties could also be beneficial in areas beyond scRNAseq, such as *in vitro* RNA synthesis and structural RNA applications.

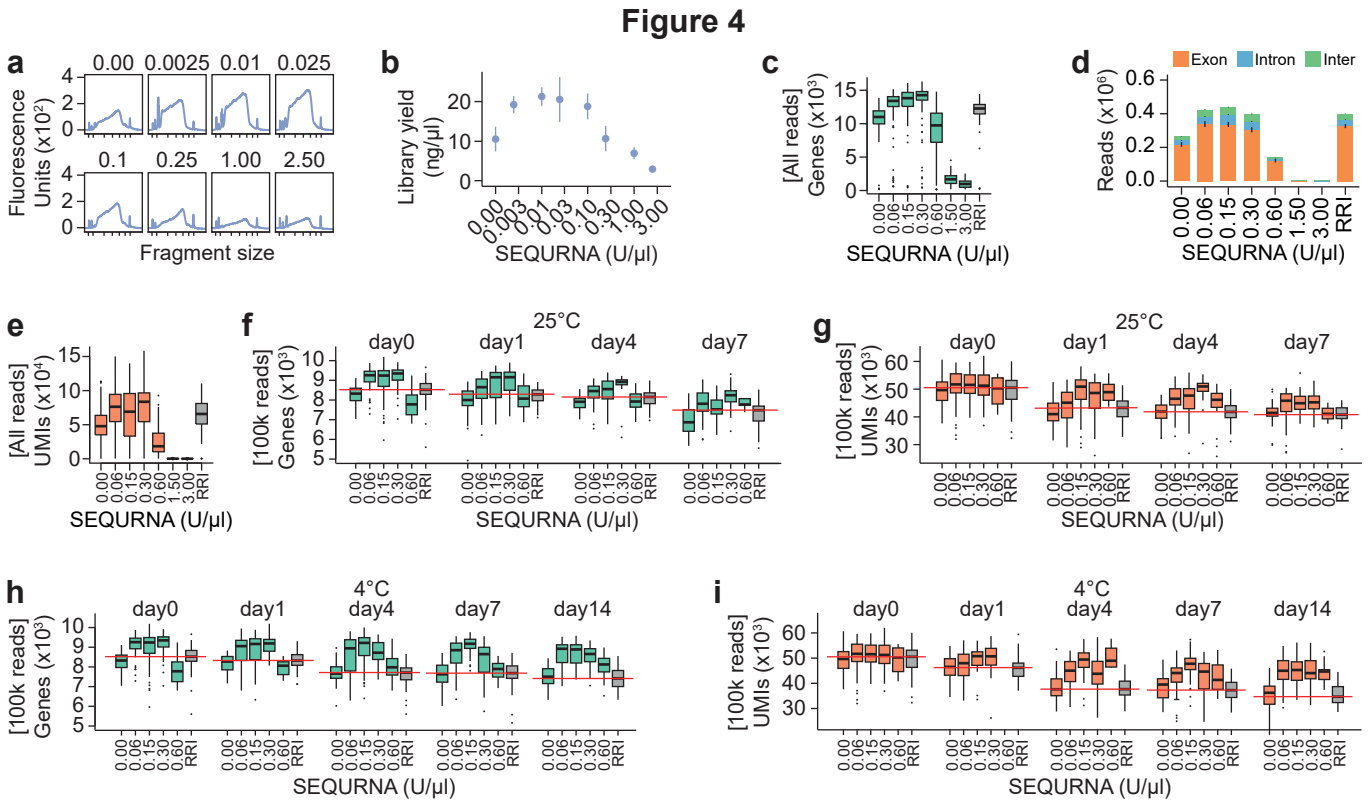


Figure 4. Characteristics of Smart-seq3xpress single-cell libraries after long-term storage in SEQURNA lysis buffer.

a. Bioanalyzer traces of Tn5-tagmented and amplified sequencing libraries of HEK293FT single-cell cDNA varying SEQURNA concentration in the tagmentation reaction mixture. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **b.** Quantified library yield of Tn5-tagmented and amplified sequencing libraries. Data shown as mean (dot) and standard error (lines) of n = 9 per condition (72 samples analyzed). **c.** Box plot of number of genes detected in Smart-seq3xpress (SS3xpress) libraries of HEK293FT cells collected in SS3xpress lysis buffer containing no inhibitor (0 U/µl), SEQURNA (0.06, 0.15, 0.3, 0.6, 1.5, or 3.0 U/µL), or standard SS3xpress using RRI. Data shown as median, interquartile range (IQR), and 1.5x IQR for n = 88-96 cells per condition and n total= 757 libraries (Day0 only). **d.** Stacked bar plot of number of reads mapping to exonic, intronic, or intergenic regions of the human genome for single-cell HEK293FT SS3xpress libraries. Data shown as mean with lines indicating standard error for n = 88-96 per condition (Day0 only). **e.** Box plots of number of unique molecular identifiers (UMIs) detected in single-cell HEK293FT SS3xpress libraries for n = 88-96 cells per condition. (**f**) Box plot of number of genes detected in single-cell HEK293FT SS3xpress libraries upon varying RNase inhibitor conditions in the lysis buffer and storing the cells for 0, 1, 4, or 7 days at 25°C. Red lines indicate the medians for standard RRI conditions. Data down sampled to 100,000 reads per cell. n = 57-93 cells per condition for Day0 and n = 6-48 cells per condition for Days1-7. n total= 1220 down sampled libraries. (**g**) Box plot of number of UMIs detected corresponding to the cells in (**f**). (**h**) Box plot of number of genes detected in single-cell HEK293FT SS3xpress libraries upon varying RNase inhibitor conditions in the lysis buffer and storing the cells for 0, 1, 4, 7, or 14 days at 4°C. n total= 1448 down sampled libraries. Red lines indicate the medians for standard RRI conditions. Data down sampled to 100,000 reads per cell. n = 57-96 cells per condition for Day0 and n = 11-48 cells per condition for Days1-7 (**i**) Box plot of number of UMIs detected corresponding to the cells in (**h**). A total of 3456 cells were sequenced (pre down sampling).

An obvious strength of a chemically synthesized RNase inhibitor is that production- and storage-related batch effects can be kept at a minimum, while protein-based RRIs are invertedly prone to degeneration and lot variation inherent to protein expression, purification, and storage over time. Moreover, production-, storage-, and transportation-related costs can be reduced.

Thermostability and direct compatibility with Tn5 tagmentation further enables multiple use cases in single-cell multiomics techniques. Although we used a second-generation sequencing-based readout in the current study (short-read sequencing), we anticipate the utility of synthetic RNase inhibitors in third-generation long-read direct sequencing of RNA and cDNA.

We moreover envision that synthetic RNase inhibitors will be exceedingly beneficial in applications requiring large liquid volumes and high inhibitor consumption, which could facilitate *in situ* RNA sequencing in tissues or whole organisms.

In summary, the replacement of RRIs with synthetic thermostable RNAse inhibitors represents a milestone in single-cell transcriptomics development.

Author contributions

JCN, AL, and BR conceived the study. JCN performed experiments. MHJ performed Smart-seq3xpress experiments. JCN, AL, MHJ, RS, and BR analysed the data. JCN, AL and BR prepared figures. JCN, AL, and BR wrote the paper. BR and RS provided resources. BR supervised the work. All authors read and edited the manuscript.

Acknowledgements

AL is supported by the Swedish Society for Medical Research (PD20-0217). BR is supported by the Swedish Society for Medical Research (CG-22-0260-H-02), the Knut & Allice Wallenberg Foundation (2021.0142 and 2022.0146), the Swedish Research Council (2022-01620), and KI SFO StratRegen 2021. We thank members from Reinius' lab for input on the paper.

Declaration of interests

JCN, AL and BR have filed patent applications on synthetic RNase inhibitors in scRNAseq and other applications and are co-founders of SEQURNA AB, making available the thermostable RNase inhibitor mix herein described as a quality-controlled kit. JCN, AL, MHJ, RS and BR are shareholders of SEQURNA AB.

Data Availability

Sequencing data generated in this study has been deposited to ArrayExpress (accession E-MTAB-13873) and computational code at github.com/reiniuslab/SEQURNA. All data and code will be made available upon publication of the peer-reviewed version of this manuscript.

Methods

Description of the synthetic RNase inhibitor

SEQURNA consists of a proprietary mix of synthetic molecules that interact with RNases in a non-linear fashion. As such, we describe SEQURNA concentrations in terms of mass Units/µl (herein abbreviated U/µl) which is not equivalent to the international unit for enzyme (IU) catalytic activity of RRIs. Fresh tubes of RRI (Cat. 2313B TaKaRa) were used in all comparisons in this study.

Animals and cells

C57BL/6J (B6) mice were crossed with M. castaneous (CAST/EiJ) mice to produce B6/CAST hybrid mice. Mice were housed in specific pathogen-free at Comparative Medicine Biomedicum (KM-B) according to Swedish national regulations for laboratory animal work food and water ad libitum, cage enrichment, and 12 hours light and dark cycles (ethical permit 17956-2018, Jordbruksverket). Primary fibroblasts were derived from adult CAST/EiJ × C57BL/6J mice by skinning, mincing, and culturing tail explants in fibroblast medium (DMEM/10% FBS) in 5% CO2 and 37°C. HEK293FT cells were cultured and expanded in standard medium (DMEM/10% FBS) in 5% CO2 and 37°C, and 10 million cells were collected and resuspended in 0.5% BSA in PBS for cell sorting. Mouse liver and spleen were collected from

purebred C57BL/6J mice. For liver and spleen collection, 25 mL of warmed perfusion buffer (140 mM NaCl, 6.7 mM KCl, 9.6 mM HEPES, 6 mM NaOH) was slowly injected into the left ventricle after a small incision in the right ventricle. Both organs were collected on ice in perfusion buffer. The spleen was cut into pieces, mashed with a syringe plunger and filtered through a 70-µm cell strainer with the addition of 2% FBS in PBS. The liver was cut into pieces and treated with 1 mg/ml collagenase in perfusion buffer with 3 mM $CaCl_2$ and 1mM $MgCl_2$ at 37 °C with shaking for 20 minutes. Liver cells were mashed through a 70-µm cell strainer with the syringe plunger, rinsing the strainer with 10% FBS in PBS. Both strained tissue types were pelleted, reconstituted in 1x RBC lysis buffer for 5 minutes at room temperature, and centrifuged at 500g for 5 minutes to remove red blood cells. Spleen cells were washed with 0.5% BSA in PBS, reconstituted in 0.5% BSA in PBS, and filtered through a 40-µm cell strainer. Liver cells were washed 2x with PBS, incubated with 2 ml TrypLE express for 10 minutes at 37 °C with shaking, diluted to 20 ml with 10% FBS in PBS, pelleted, reconstituted in 0.5% BSA in PBS, and filtered through a 40-µm cell strainer. The cell suspensions in 0.5% BSA in PBS were then subjected to single-cell sorting by FACS.

Generation of Smart-seq2 and Smart-seq3 sequencing libraries

Bulk RNA was extracted from cultured mouse tail tip fibroblasts using TRIzol (Invitrogen). For all bulk RNAseq experiments 100 pg of RNA was used as input, if is not otherwise specified, and was added to 96-well plates containing either Smart-seq2 or Smart-seq3 standard lysis buffer or lysis buffer with varying concentrations of SEQURNA. For Smart-seq2 and Smart-seq3 single-cell experiments, cultured HEK293FT, spleen, or liver cells in 0.5% BSA in PBS were sorted into 96 well plates containing lysis buffer using an SH800S Cell Sorter (Sony). Single-cell sorted plates were briefly centrifuged and kept at -80°C until they were further processed. Plates containing lysis buffer and input material were then processed according to the Smart-seq2 and Smart-seq3 protocols[5,9], except without addition of recombinant inhibitor in the first-strand reaction mix in reactions where SEQURNA was used.

Smart-seq2 cDNA library generation: Smart-seq2 lysis buffer composition was 0.08% Triton X-100, 2.2mM dNTP/each, 2.2mM Smart-seq2 oligo-dT (5′-AAGCAGTGGTATCAACGCAGAGTACT30VN-3′), and 4 U recombinant inhibitor (Cat. 2313B, TaKaRa) or SEQURNA thermostable RNase inhibitor (Cat. SQ00201, SEQURNA) of the concentration specified for each condition; total buffer volume 4.5 µl. Cells were lysed and RNA denatured at 72°C for 3 min in a Bioer Life ECO thermocycler and placed on ice prior to first-strand synthesis. The following reverse transcriptase reaction contained 1x Superscript II buffer, 5mM DTT, 1M betaine, 10 mM $MgCl_2$, 1µM Smart-seq2 TSO (5′-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3′), 100 U Superscript II, and 10 U recombinant inhibitor (TaKaRa) for samples containing biological inhibitor; total reaction volume 10 µl. RT thermocycles were 42°C for 90 min, followed by 10 cycles of 42°C for 2 min and 50°C for 2 min. The following cDNA amplification reaction contained 1x KAPA HiFi HotStart Ready Mix and 80nM ISPCR primers (5′-AAGCAGTGGTATCAACGCAGAGT-3′); total reaction volume 25 µl. Thermocycles for Smart-seq2 cDNA amplification were 98°C for 3 min, followed by either 18 (bulk), 20 (HEK), 21 (liver) or 22 (spleen) cycles of 98°C for 20 sec, 67°C for 15 sec, and 72°C for 6 min, followed by a final incubation at 72°C for 5 min.

Smart-seq3 cDNA library generation: Smart-seq3 lysis buffer composition was 0.1% Triton X-100, 10% PEG 8000, 1mM dNTP/each, and 1µM Smart-seq3 oligo-dT (5′-biotin-ACGAGCATCAGCAGCATACGA T30VN-3′), and 1.2 U recombinant inhibitor (Cat. 2313B, TaKaRa) or SEQURNA Thermostable RNase inhibitor (Cat. SQ00201, SEQURNA) of the concentration specified for each condition; total buffer volume 2 µl. Cells were lysed and RNA denatured at 72°C for 10 min and placed on ice prior to first-strand synthesis. The following reverse transcriptase reaction contained 50 mM Tris-HCl (pH 8.3), 30mM NaCl, 1mM GTP, 8mM DTT, 2.5mM MgCl2, 2µM Smart-seq3 TSO (5'-biotin-AGAGACAGATTGCGCAATGNNNNNNNNrGrGrG-3'), 8 U Maxima H-minus RT enzyme, and

2 U recombinant inhibitor (TaKaRa) for samples containing biological inhibitor; total reaction volume 4 µl. RT thermocycles were 42°C for 90 minutes, followed by 10 cycles of 42°C for 2 min and 50°C for 2 min. The following cDNA amplification reaction contained 1x KAPA HiFi buffer, 300nM dNTP/each, 500nM MgCl2, 500nM forward primer (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTGCGCAA*T*G-3′), 100nM reverse primer (5′-ACGAGCATCAGCAGCATAC*G*A-3′), and 0.2 U KAPA polymerase in a total reaction volume of 10 µl. Thermocycles for Smart-seq3 cDNA amplification were 98°C for 3 min, then either 20 (bulk) or 22 (single-cell) cycles of 98°C for 20 sec, 65°C for 30 sec, and 72°C for 6 min, followed by a final incubation at 72°C for 5 min.

Amplified cDNA was bead purified (AMPure XP, Beckman) at a ratio of 0.8:1 beads:cDNA (20 µl beads:25 µl cDNA for Smart-seq2 and 8 µl beads:10 µl cDNA for Smart-seq3) and inspected on a Bioanalyzer 2100 (High Sensitivity DNA kit, Agilent).

For Smart-seq2, the tagmentation reaction contained 2xTAPS-Mg buffer, 10% PEG 8000, 0.3nM of a custom Tn5[11], and 1ng of cDNA in 20 µl total, was incubated for 8 minutes at 55°C, and Tn5 was stripped from the DNA with 3.5 µl of 0.2% SDS, incubated at room temperature for 5 minutes. To amplify Smart-seq2 libraries, a reaction mix containing the tagmented cDNA, 0.75 µl of 1 µM pooled dual-index primers, 1x Kapa HiFi PCR buffer, 300µM dNTP/each, and 1 U Kapa HiFi polymerase in 50 µl total reaction was run in a thermocycler for 3 minutes at 72°C, 95°C 30 seconds, then 10 cycles of 95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, then 72°C for 5 minutes.

For Smart-seq3, the tagmentation reaction contained 10mM Tris-HCl pH 7.5, 5mM MgCl2, 5% dimethylformamide, and 0.08 µl (0.12 µl for single cell) Tn5 (Nextera Amplicon Tagment Mix, Illumina) and 100pg of cDNA in 2µl total, was incubated for 10 minutes at 55°C, and Tn5 was stripped from the DNA with 0.5 µl of 0.2% SDS (0.4 µl of 0.2% SDS for single-cell). To amplify Smart-seq3 libraries, a reaction mix containing the tagmented cDNA, 0.75 µl of 1 µM pooled dual index primers, 1x Phusion HiFi PCR buffer, 200µM dNTP/each, and 0.1 U Phusion HiFi polymerase in 7 µl total was run on a thermocycler at 72°C, 98°C for 3 minutes, then 12 cycles of 98°C for 10 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, then 72°C for 5 minutes.

Indexed libraries were bead purified (AMPure XP, Beckman) at a ratio of 0.6:1 beads:cDNA (30 µl beads:50 µl library for Smart-seq2 and 4.2 µl beads:7 µl library for Smart-seq3). All Smart-seq2 libraries were sequenced as 74bp, single-end reads, and all Smart seq3 libraries were sequenced as 74bp, paired-end reads.

Bioanalyzer DNA yield calculations

All bioanalyzer traces were exported in raw format as a csv file. For all histograms, the average and standard error of the area-under-curve (DNA yield) was calculated for each sample. Percent primer dimer was calculated as (primer-dimer yield)/(total library yield). Number of replicas used for quantified cDNA and primer-dimer yield can be found in **Supplementary Table 1**.

Read alignment and genes and UMIs detected

Read files were downsampled to 100k and aligned using STAR[12] against the Gencode GRm38 (mouse) or GRCh38 (human) genome assembly. STAR aligner determined the accuracy of mapping for each sample with percent uniquely mapped reads and percent unmapped reads. STAR was also used to calculate the rate of read mismatch to the genome and length of insertions and deletions within the read relative to the genome. MultiQC's output of STAR generated a summary of all STAR QC parameters for all samples. Data for both number of genes and UMIs detected was generated using the zUMIs pipeline[13]. Briefly, downsampled loom files that contained read coverage of exons only were processed in R for gene count (readcount.exon) and UMI count (umicount.exon). The number of genes expressed for each sample was quantified as sum of every gene with ≥ 1 read fragment. The number of UMIs detected for each sample was quantified as the sum of all UMI-containing fragments. Number of samples used for quality control analysis after downsampling (>100k reads) can be found in **Supplementary Table 1**.

Base quality along read length

BBMap's tool mhist was used to plot the frequency that a read's base position matched or contained a substitution/insertion/deletion relative to the genome. Output files for each type were edited to and merged as a usable matrix for downstream processing in R.

Genomic and gene body mapping

Sorted bam files generated by STAR were used in more in-depth mapping quantifications. The fraction of reads that mapped to exonic, intronic, or intergenic regions in the genome were obtained with Qualimap's RNA-seq QC. MultiQC's output of Qualimap generated a summary of all parameters for all samples. Gene body coverage plots were generated using computeMatrix and plotProfile via deepTools.

Dimensionality reduction, cell cycle scoring, and clustering

For non-subsampled Smart-seq2 HEK data, cells with 3 MADs lower log10(genes detected) or log10(readcounts) than global median were excluded and readcounts were log-normalised then centered and scaled using Seurat[14] (v4.3; NormalizeData, ScaleData). Next, cells were scored for cell cycle using Seurat (CellCycleScoring, cc.genes.updated.2019) and PCA was performed using cell cycle-related genes used for cell cycle scores. For non-subsampled Smart-seq2 spleen and liver data, cells were filtered and normalized as described above, and 2,000 variable features were selected using Seurat (FindVariableFeatures, selection.method="vst"). PCA was performed using variable features and UMAP and nearest-neighbor detection was performed for the first 50 PCs using k=30 nearest neighbors (RunUMAP, dims = 1:50, n.neighbors=30; FindNeighbors, dims = 1:50, k.param = 30). Clusters were identified using Seurat (FindClusters, resolution = 0.3) and cluster identities were manually annotated based on gene expression. Number of samples passing quality control (< 3 MAD below the median genes detected or readcounts) used for dimensionality reduction analyses can be found in **Supplementary Table 1**.

Data visualization

All data were plotted using ggplot2 or ComplexHeatmap in R. (T-tests were performed using pairwise_t_test from the package tidytests).

PCR for primer dimer and non-specific band assays

The primer dimer assay was designed to produce a 164 bp amplicon from a Cas9-eGFP plasmid. The final PCR reaction mix contained 1x Hifi Kapa Mix, 1µM of the fwd/rev primer pool, and 1ng of plasmid. PCR reaction conditions were 95°C for 3 min, then 25 cycles of 98°C for 20 sec, 65°C for 15 sec, and 72°C for 30 sec, followed by a final extension at 72°C for 1 min. The non-specific band assays were designed to produce amplicons between 120 and 140 bp from Tuba1a in the human genome. The final PCR reaction mix for the non-specific band assay contained 1x Kapa Hifi Mix, 1µM of the fwd/rev primer pool, and 20ng of HEK293FT gDNA. PCR thermocycles were 95°C for 3 min, then 25 cycles of 98°C for 20 sec, 57°C for 15 sec, and 72°C for 30 sec, followed by a final extension at 72°C for 1 min.

Melt curve analysis

Purified eGFP amplicons (164 bp) were used as a dsDNA template. The reaction mixture contained 1x SYBR Mix, 500ng of template, and varying concentrations of SEQURNA (n=6 for each sample). The assay conditions were: 95°C for 15 sec to denature, 40°C for 1 min to anneal, and a temperature gradient from 40 to 95°C at a rate for 0.1° per second. Melt curve was performed with StepOne Plus Real-Time machine (Applied Biosystems). Results were exported to csv files using the Applied Biosystems Analysis Software.

Smart-seq2 storage-time experiment

HEK293FT cells were FACS-sorted into Smart-seq2 lysis buffer (0.08% Triton X-100, 2.2mM dNTP/each, and 2.2mM Smart seq2 oligo-dT) containing 0, 6, 9, or 12 U/µl of SEQURNA in 96 well plates, and were subsequently stored at either 25°C, 4°C, or -80°C. After 1, 4, 7, or 14 days, two replicas of each plate storage condition were processed for cDNA following the Smart-seq2 protocol (omitting the biological inhibitor in the reverse transcriptase step as described in a previous section) with 20 cycles of PCR

amplification. An additional two plates were immediately processed into cDNA following the Smart-seq2 protocol after sorting as a control (day 0). After cDNA library amplification, samples were bead purified (AMPure XP, Beckman Coulter) and run on High Sensitivity DNA chips using a Bioanalyzer 2100 (Agilent) to assess cDNA quality.

Smart-seq3xpress storage-time experiment and library preparation

Smart-seq3xpress libraries were performed as previously described[10] with some modifications. In brief, cells were sorted into 384 well plates each containing 3 µl Vapor-Lock (Qiagen) in all wells and eight different conditions of 0.3 µl lysis buffer consisting of 0.125µM OligodT30VN (5'-Biotin-ACGAGCATCAGCAGCATACGAT30VN-3' ; IDT) adjusted to RT volume, 0.5mM dNTPs/each adjusted to RT volume, 0.1% Triton X-100, 5% PEG8000 adjusted to RT volume, and the indicated type and amount of RNase Inhibitor (no RNAse inhibitor, 0.003 µl RNase Inhibitor (40 U/µl, Cat. 2313B, TaKaRa), SEQURNA Thermostable RNase inhibitor (Cat. SQ00201, SEQURNA); 0.06 U/µl, 0.15 U/µl, 0.3 U/µl, 0.6 U/µl, 1.5 U/µl, 3.0 U/µl). After cell sorting plates were briefly centrifuged before put to -80°C immediately after sorting (day 0). To test the effect of temperature and time on single cell RNA stability for each of the eight lysis conditions, 384 well plates containing sorted cells in lysis buffer were left at either room temperature (25 °C) for 1, 4, or 7 days or in fridge (4°C) for 1, 4, 7, or 14 days, before commencing library preparation. To serve as control (day 0), two plates were extracted immediately from -80°C immediately before library preparation.

Before RT, plates were denatured at 72°C for 10 min followed by addition of 0.1 µl of RT mix; 25mM Tris-HCL pH 8.4 (Fischer Scientific), 30mM NaCl (Ambion), 1mM GTP (Thermo Fisher Scientific), 2.5mM $MgCl_2$ (Ambion), 8mM DTT (Thermo Fisher Scientific), 0.75uM Template Switching Oligo (TSO) (5′-Biotin-AGAGACAGATTGCGCAATGNNNNNNNNW WrGrGrG-3′; IDT), 0.25 U/ µl RRI (Cat. 2313B, TaKaRa), and 2 U/µl of Maxima H Minus reverse transcriptase (Cat. EP0752, Thermo Fisher Scientific). RRI was excluded from the RT mix (replaced with water) in case of SEQURNA samples. Plates were quickly centrifuged after dispensing to ensure merge of lysis and RT volumes underneath the Vapor-lock overlay and incubated at 42 °C for 90 min, followed by ten cycles of 50 °C for 2 min and 42 °C for 2 min. After RT, 0.6 µl PCR mix was dispensed to each well containing the following: 1× SeqAmp PCR buffer (Takara Bio), 0.025 U µl−1 of SeqAmp polymerase (Takara Bio) and 0.5 µM Smartseq3 forward (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGATTGCGCAATG-3′; IDT) and reverse primer (5′-ACGAGCATCAGCAGCATACGA-3′; IDT). Plates were quickly spun down before being incubated as follows: 1 min at 95 °C for initial denaturation, 12 cycles of 10 sec at 98 °C, 30 sec at 65 °C and 4 min at 68 °C. Final elongation was performed for 10 min at 72 °C.

After PCR, pre-amplified libraries were diluted with 9 µl H2O, before transferring 1 µl of diluted cDNA from each well into a new 384 well plate. Tagmentation was performed by adding 1 µl of tagmentation mix; 1x tagmentation buffer (10 mM Tris pH 7.5, 5 mM MgCl2, 5% DMF), 0.003 µl Tagmentation DNA Enzyme 1 (TDE1; Illumina DNA sample preparation kit) to the 1µl of diluted cDNA per well. At this step the tagmentation mix contains SEQURNA RNase inhibitor at a concentration of 0.015x what was in the lysis buffer. Plates were incubated for 10min at 55°C before the reaction was stopped by the addition of 0.5 µl 0.2% SDS to each well. Index PCR was carried out after the addition of 3.5µl custom Nextera Index primers (0.5µM) by dispensing 2 µl of PCR mix; 1× Phusion Buffer (Thermo Fisher Scientific), 0.01 U µl−1 of Phusion DNA polymerase (Thermo Fisher Scientific), 0.025% Tween-20, 0.2 mM dNTP each. PCR was performed out at 3 minutes at 72 °C; 30 seconds at 95 °C; 12 cycles of (10 seconds at 95 °C; 30 seconds at 55 °C; 1 minute at 72 °C); and 5 minutes at 72 °C. Each indexed library plate was pooled by spinning out gently using a 300-ml robotic reservoir (Nalgene) fitted with a custom scaffold by pulsing the centrifuge to < 200g. The pooled libraries were afterwards purified with custom carboxylated magnetic beads in 22% PEG solution at a ratio of 1 sample to 0.7 beads.

Sequencing of Smart-seq3xpress libraries

Smart-seq3xpress libraries were sequenced on a MGI DNBSEQ G400RS platform. Prior to sequencing on MGI platform ready circular ssDNA libraries were generated using the MGIEasy Universal Library Conversion Kit (MGI). Adapter conversion PCR was carried out on 50ng of final pooled library for 5 cycles, following circularization of 1pmol dsDNA according to manufacturer's protocol. DNA nanoballs (DNBs) were created from 80 fmol of circular ssDNA library pools using a custom rolling-circle amplification primer (5′-TCGCCGTATCATTCAAGCAGAAGACG-3′, IDT). DNBs were sequenced 100 bases paired end (PE100) using custom sequencing primers (Read 1: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′; MDA: 5′-CGTATGCCGTCTTCTGCTTGAATGATACGGCGAC-3′, Read 2: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′; i7 index: 5′-CCGTATCATTCAAGCAGAAGACGGCATACGAGAT-3′; i5 index: 5′-CTGTCTCTTATACACATCTGACGCTGCCGACGA-3′).

Data preprocessing of Smart-seq3xpress libraries

Raw FASTQ files were processed with zUMIs[13] 2.9.7 pipeline. UMI-containing reads were identified by the (ATTGCGCAATG) pattern allowing up to two mismatches and reads were filtered for low quality UMIs (3 bases < phred 20) and index barcodes (4 bases < phred 20), before mapped to the human genome (hg38) using STAR5 version 2.7.3. Read counts, detected genes, and umicounts were calculated using Ensembl gene annotations (GRCh38.95). Down sampled data (100k reads) was generated via zUMIs preprocessing. Number of samples used for quality control analysis after downsampling (>100k reads) can be found in **Supplementary Table 1**.

Inhibitor-spiked tagmentation

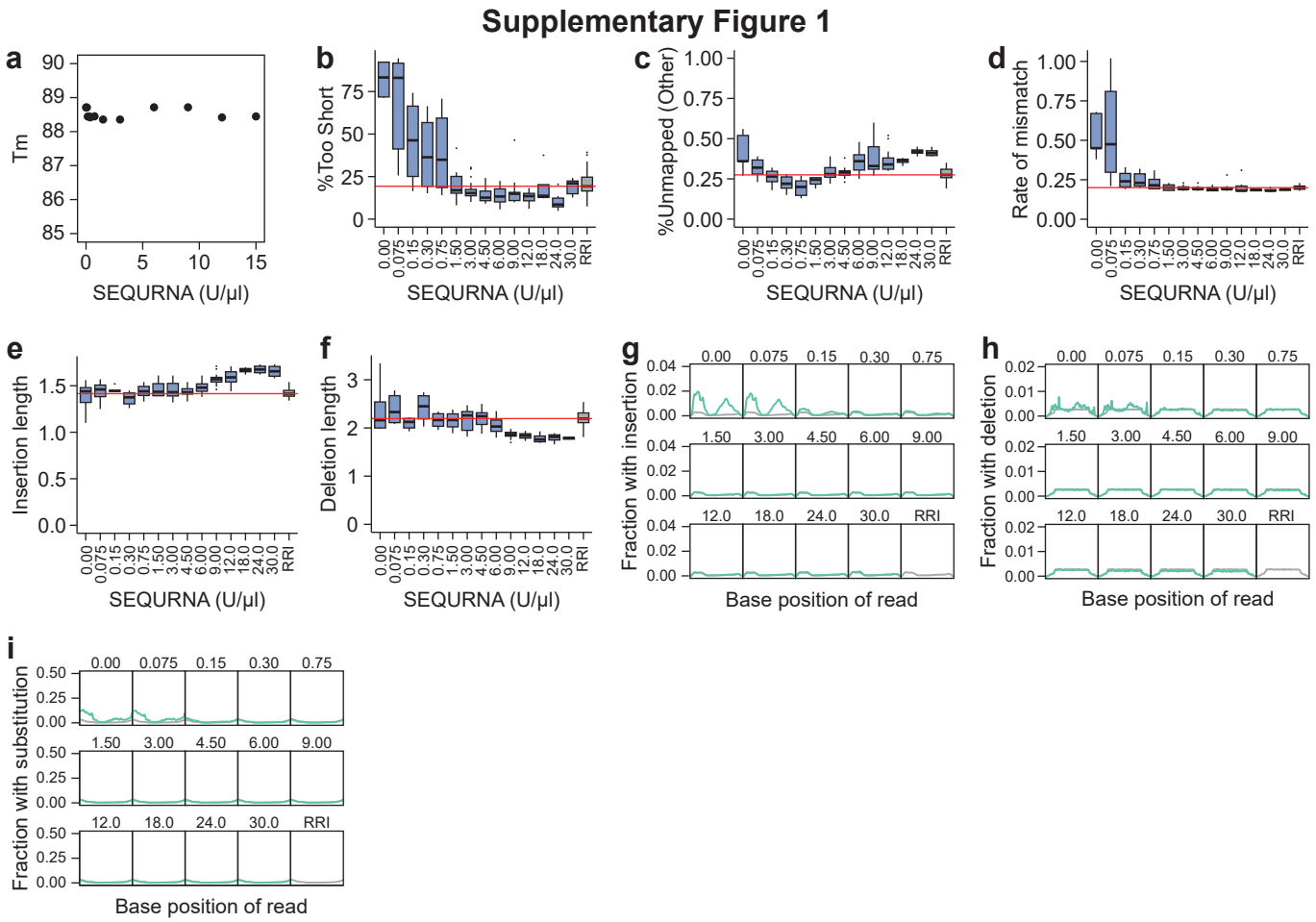
Bulk HEK RNA (100pg) was used to generate amplified cDNA with the standard Smart-seq2 protocol (18 cycles, full 96-well plate) replacing RRI in the lysis buffer with SEQURNA (1.2 U/µl). All wells were pooled and bead purified (AMPure XP, Beckman) at a ratio of 0.8:1 beads:cDNA. Purified, amplified cDNA was diluted to a 1 ng/µl stock to use for tagmentation.

The tagmentation reaction contained 2xTAPS-Mg buffer, 10% PEG 8000, 0.3nM Tn5, and 2ng of cDNA. SEQURNA was added to the reactions at final concentrations of 0.0025, 0.01, 0.025, 0.1, 0.25, 1, or 2.5 U/µl (or no inhibitor) in 20µl total. The reaction was incubated for 8 minutes at 55°C, and Tn5 was stripped from the DNA with 3.5 µl of 0.2% SDS, incubated at room temperature for 5 minutes. Tagmented libraries were amplified in a reaction mix containing 1µl of pooled custom dual-index primers, 1x Kapa HiFi PCR buffer, 300uM dNTP/each, and 1 U Kapa HiFi polymerase in 50 µl total, and was run in a thermocycler for 3 min at 72°C, 95°C 30 sec, then 10 cycles of 95°C for 10 sec, 55°C for 30 sec, and 72°C for 30 sec, then 72°C for 5 min. Indexed libraries were bead purified (AMPure XP, Beckman) at a ratio of 1:1 beads:library (50 µl beads:50 µl library). Library concentration was quantified with the QuantiFluor ONE dsDNA System using the Varioskan LUX (Thermo Scientific) microplate reader. Library quality was visualized on a Bioanalyzer 2100 (High Sensitivity DNA kit, Agilent).

References

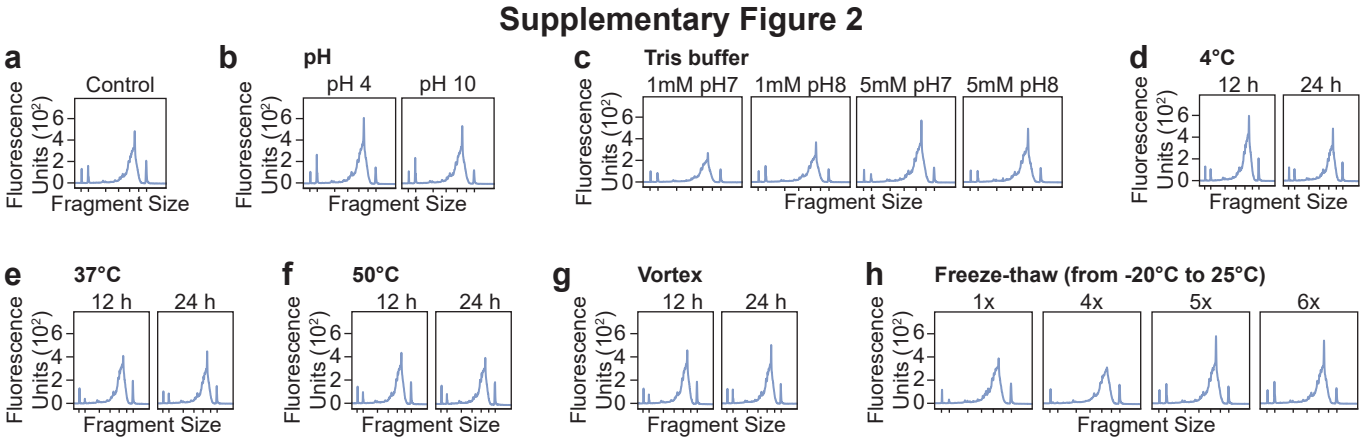
1. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* **9**, 75 (2017).
2. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol Cell* **58**, 610–620 (2015).
3. Trombetta, J. J. *et al.* Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Curr Protoc Mol Biol* **107**, 4.22.1-4.22.17 (2014).
4. Martin, B. K. *et al.* Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat Protoc* **18**, 188–207 (2023).
5. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096–1098 (2013).
6. Johnston, A. D., Lu, J., Ru, K.-L., Korbie, D. & Trau, M. PrimerROC: accurate condition-independent dimer prediction using ROC analysis. *Sci Rep* **9**, 209 (2019).

7. Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).
8. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
9. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* **38**, 708–714 (2020).
10. Hagemann-Jensen, M., Ziegenhain, C. & Sandberg, R. Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. *Nat Biotechnol* **40**, 1452–1457 (2022).
11. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
12. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
13. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *GigaScience* **7**, (2018).
14. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).



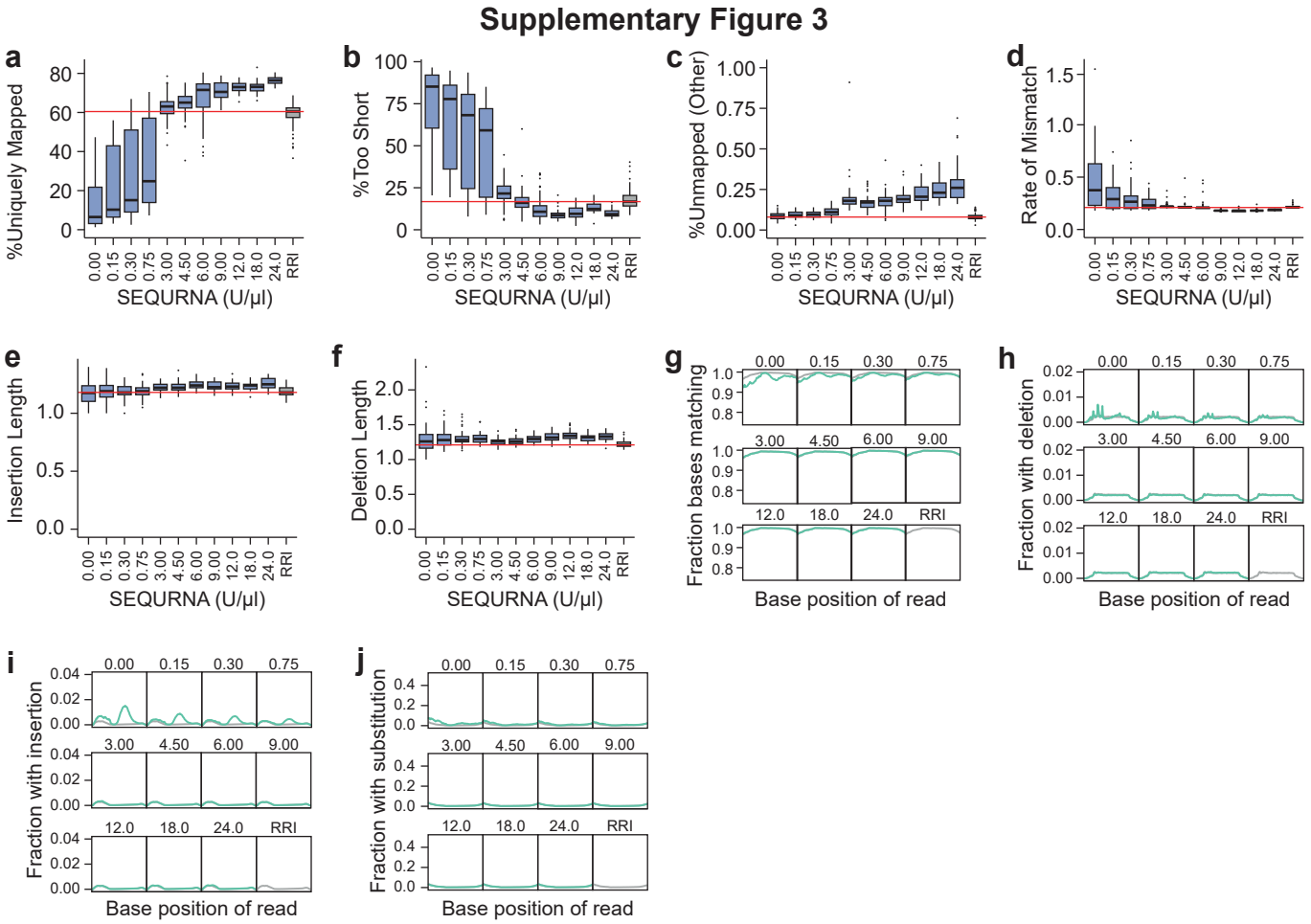
Supplementary Figure 1. Melting temperature and mapping statistics of Smart-seq2 mini-bulk libraries generated using SEQURNA RNase inhibitor.

a. Melting temperatures of a 164 bp DNA duplex under various concentrations of SEQURNA (0-15 U/µl) via denaturation detection using SYBR Green dye fluorescence in a quantitative PCR instrument. **b.** Box plot of percent sequencing reads too short to be mapped to the reference genome for Smart-seq2 (SS2) libraries, generated from 100 pg total mouse RNA and various concentrations of SEQURNA in the SS2 lysis buffer, or standard SS2 lysis buffer using RRI. **c.** Box plot of percent sequencing reads unmapped (other) to the reference genome for SS2 libraries. **d.** Box plot of the mismatch rate in reads of SS2 bulk libraries. **e.** Box plot of the average insertion length of SS2 libraries. **f.** Box plot of the average deletion length in reads of SS2 libraries. **g.** Line plot of fraction of bases along sequencing reads with insertion in SS2 libraries. **h.** Line plot of fraction of bases along sequencing reads with deletion in SS2 libraries. **i.** Line plot of fraction of bases along sequencing reads with substitution in SS2 libraries. **b-i.** n = 5-35 replicates per condition (total = 195).



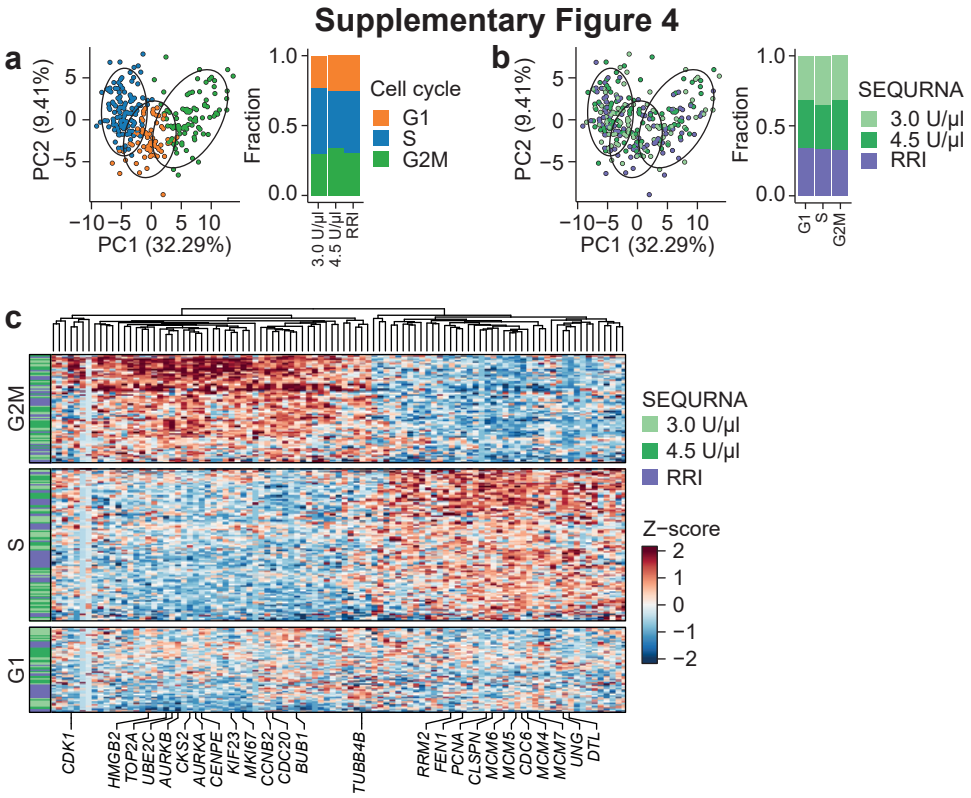
Supplementary Figure 2. Stress test of SEQURNA RNase inhibitor with subsequent Smart-seq2 library generation.

a. Bioanalyzer trace of Smart-seq2 (SS2) cDNA where the lysis buffer contained 3 U/µl SEQURNA (untreated control). **b.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of a SEQURNA which had been subjected to pH 4 (left) or pH 10 (right) using HCl and NaOH respectively before using the inhibitor in lysis buffer. **c.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of a SEQURNA kept in either in a 1mM Tris pH 7 solution, 1mM Tris pH 8 solution, 5mM Tris pH 7 solution, or 5mM Tris pH 8 solution (left to right). **d.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of SEQURNA stock that had been stored at 4°C for 12 hours (left) or 24 hours (right). **e.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of SEQURNA stock that had been stored at 37°C for 12 hours (left) or 24 hours (right). **f.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of SEQURNA stock that had been stored at 50°C for 12 hours (left) or 24 hours (right). **g.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of SEQURNA stock that had been vortexed for 12 hours (left) or 24 hours (right). **h.** Bioanalyzer traces of SS2 cDNA where the lysis buffer contained 3 U/µl of SEQURNA stock that had been freeze-thawed 1, 4, 5, or 6 times before use. **a-h.** All libraries were generated from 100 pg mouse total RNA and using 18 PCR cycles in cDNA amplification. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs.



Supplementary Figure 3. Mapping statistics of HEK293FT Smart-seq2 libraries generated using SEQURNA RNase inhibitor.

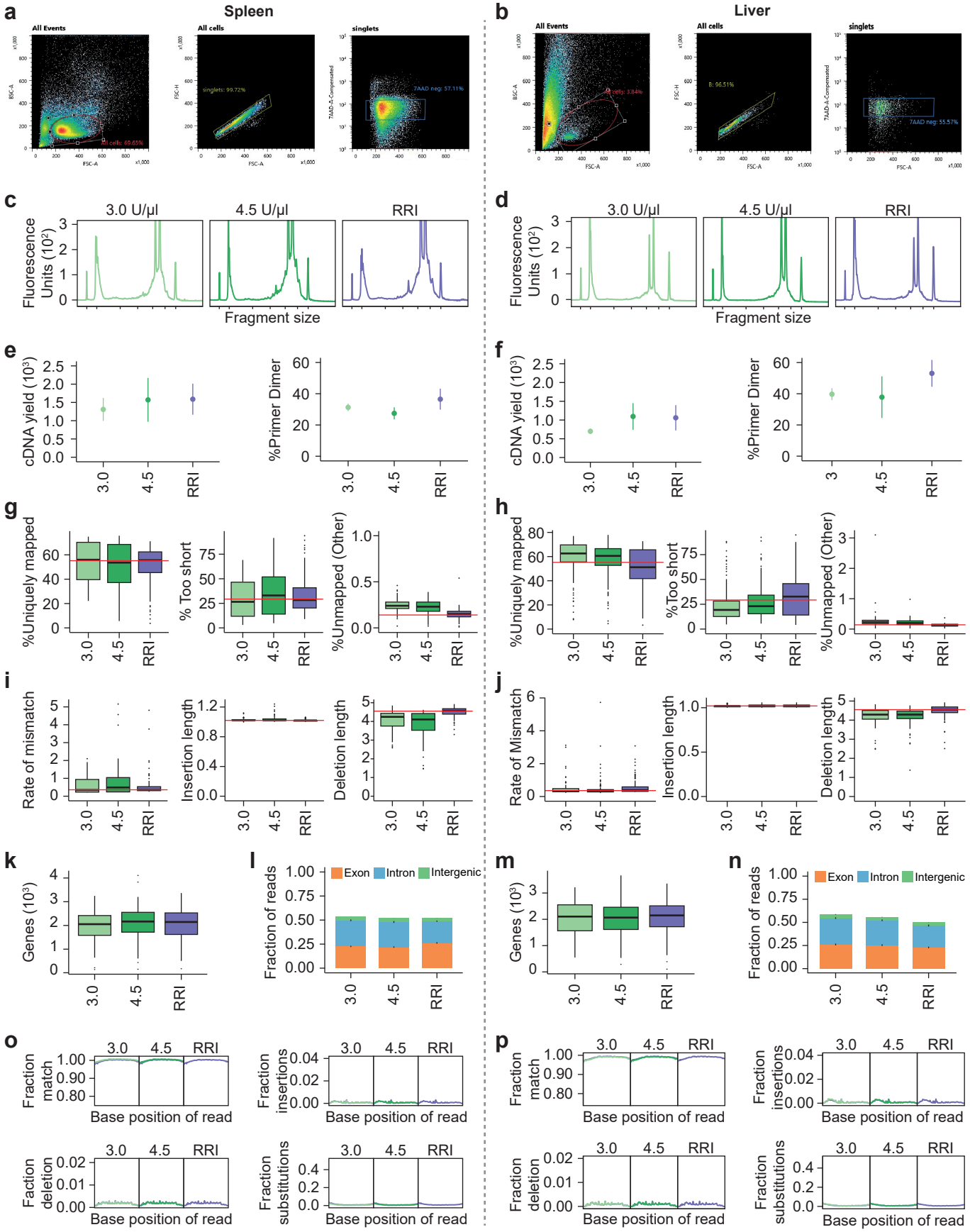
a. Box plot of percent uniquely mapped sequencing reads to the reference genome for Smart-seq2 (SS2) single-cell libraries generated from HEK293FT cells and various concentrations of SEQURNA in the SS2 lysis buffer or standard SS2 lysis buffer using RRI. n = 37-94 replicates per condition (total= 704). **b.** Box plot of percent sequencing reads too short to be mapped to the reference genome for HEK293FT SS2 libraries. **c.** Box plot of percent sequencing reads unmapped (other) to the reference genome for HEK293FT SS2 libraries. **d.** Box plot of the mismatch rate in HEK293FT SS2 bulk samples. **e.** Box plot of the average insertion length in reads of HEK293FT SS2 libraries. **f.** Box plot of the average deletion length in reads in HEK293FT SS2 libraries. **g.** Line plot of fraction of matching bases along sequencing reads of HEK293FT SS2 libraries. **h.** Line plot of fraction of bases along sequencing reads with deletion in HEK293FT SS2 libraries. **i.** Line plot of fraction of bases along sequencing reads with insertion in HEK293FT SS2 libraries. **j.** Line plot of fraction of bases along sequencing reads with substitution in HEK293FT SS2 libraries.



Supplementary Figure 4. Cell cycle analysis of HEK293FT single-cell libraries generated using SEQURNA or recombinant RNase inhibitor.

a. PCA plot for cell-cycle genes for HEK293FT Smart-seq2 (SS2) libraries generated using 3 U/µl (n = 91 cells) and 4.5 U/µl (n = 93 cells) SEQURNA in the lysis buffer, or standard SS2 lysis buffer with recombinant RNase inhibitor (RRI; n = 86 cells), colored according to cell cycle phase. Bar plot of fraction cells in cell cycle phases for each condition is shown to the right. **b.** Same as in (**a**) but coloring PCA cell data points according to RNase inhibitor condition. Bar plot of fraction of cells adhering to each RNase inhibitor condition for each cell cycle phase is show to the right. **c.** Expression-level heatmap of highly variable cell cycle genes grouped by phase for conditions 3 U/µl SEQURNA, 4.5 U/µl SEQURNA, and standard SS2 using RRI.

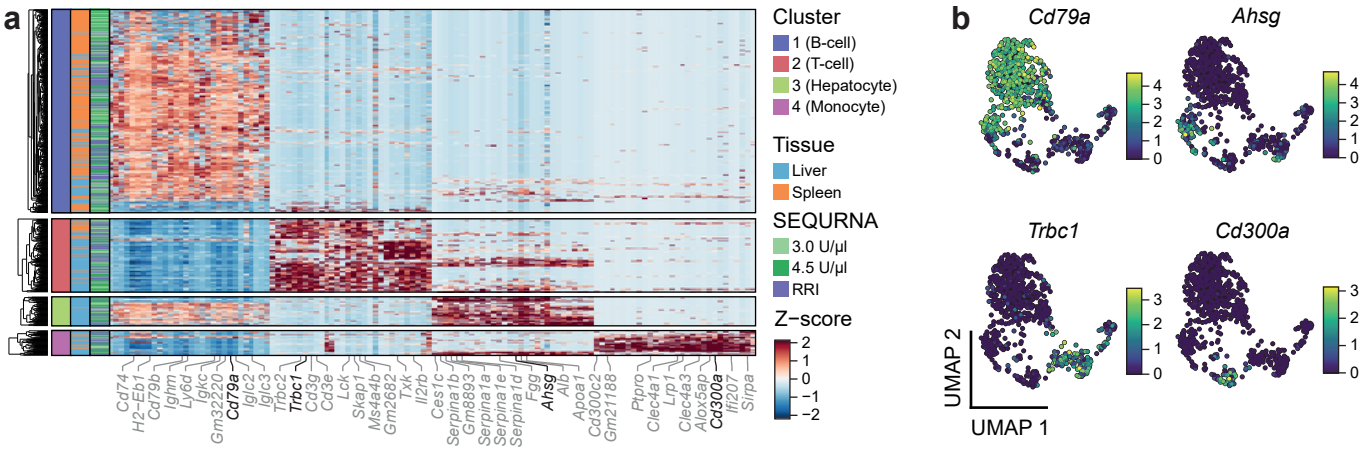
Supplementary Figure 5



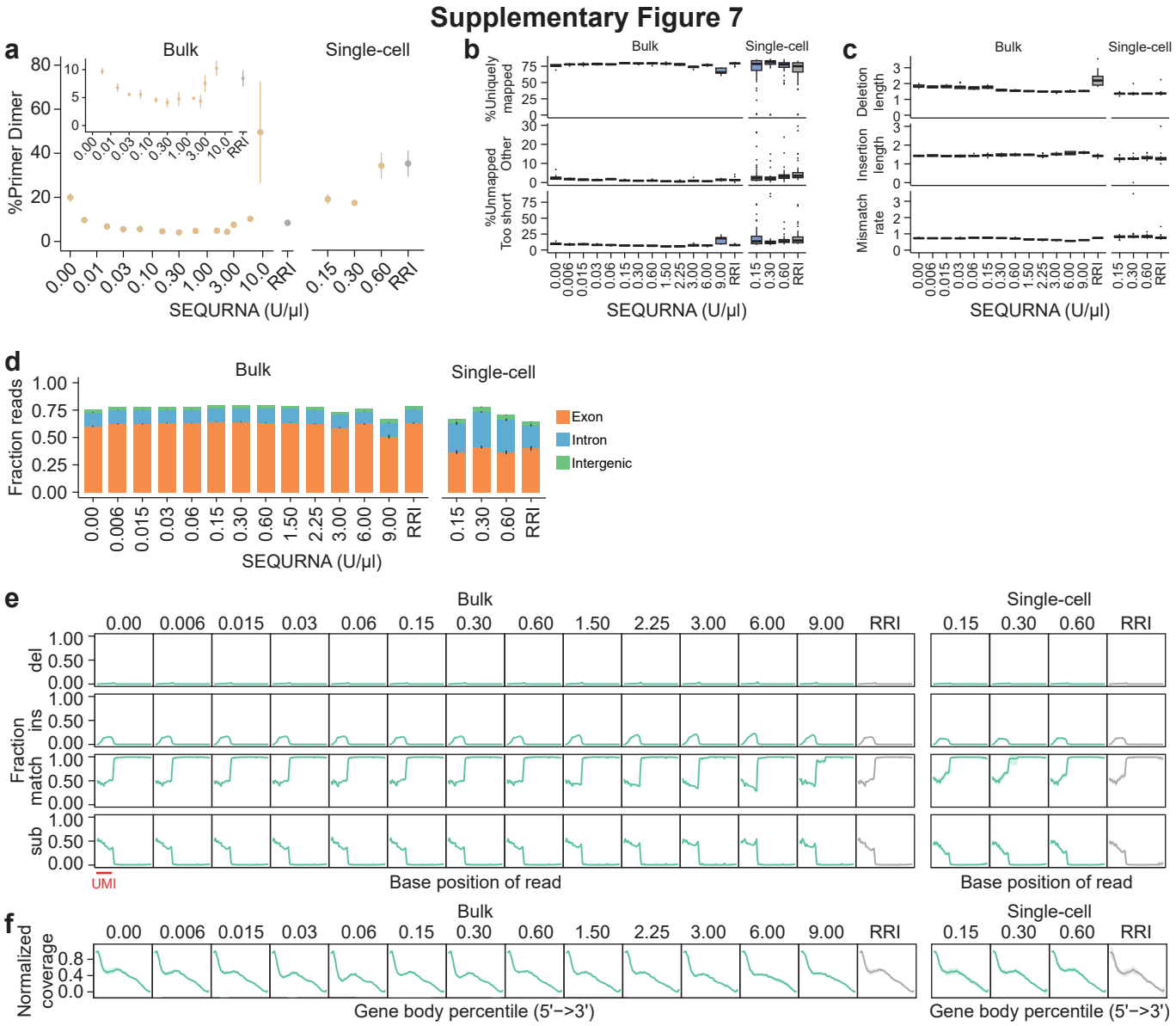
Supplementary Figure 5. FACS gate, cDNA library characteristics, and quality control of liver and spleen cells and scRNAseq libraries.

a. FACS scatter plot of spleen cells gating for lymphocyte population (left), singlets (middle), and viable cells (7AAD-neg) (right). **b.** FACS scatter plot of liver cells gating for lymphocyte population (left), singlets (middle), and viable cells (7AAD-neg) (right). **c.** Bioanalyzer traces of Smart-seq2 (SS2) cDNA libraries generated from sorted spleen cells using 3 and 4.5 U/µl SEQURNA in the lysis buffer, or standard SS2 using recombinant RNase inhibitor (RRI). **d.** Bioanalyzer traces of SS2 cDNA libraries generated from sorted liver cells using 3 and 4.5 U/µl SEQURNA in the lysis buffer, or standard SS2 using RR). Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **e.** Average cDNA yield (dot) and standard error (whiskers) of SS2 libraries from spleen cells (integration range 200–10,000 bp) (left) and average percent (dot) and range (whiskers) of primer-dimer in SS2 libraries (integration range 20–50 bp) (right). **f.** Same as in (e) but for SS2 libraries from liver. **g.** Box plots of percent uniquely mapped (left), percent sequencing reads too short to be mapped (middle), and percent sequencing reads unmapped (not due to fragments too short or too many mismatches) (right) to the reference genome for spleen cells. **h.** Same as in (g) but for SS2 libraries from liver. **i.** Box plots of rate of mismatch per base (left), average insertion length (middle), and average deletion length (right) for spleen cells. **j.** Same as in (h) but for SS2 libraries from liver. **k.** Box plot of number of genes detected for each RNase inhibitor condition for spleen cells. **l.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the mouse genome for spleen cells. **m.** Box plot of number of genes detected for each RNase inhibitor condition for liver cells. **n.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the mouse genome for liver cells. **o.** Line plot of fraction of bases along sequencing reads matching the genome (top left), with an insertion (top right), with a deletion (bottom left), and with a substitution (bottom right) indicating read quality for spleen cells. **p.** Line plot of fraction of bases along sequencing reads matching the genome (top left), with an insertion (top right), with a deletion (bottom left), and with a substitution (bottom right) indicating read quality for liver cells.

Supplementary Figure 6

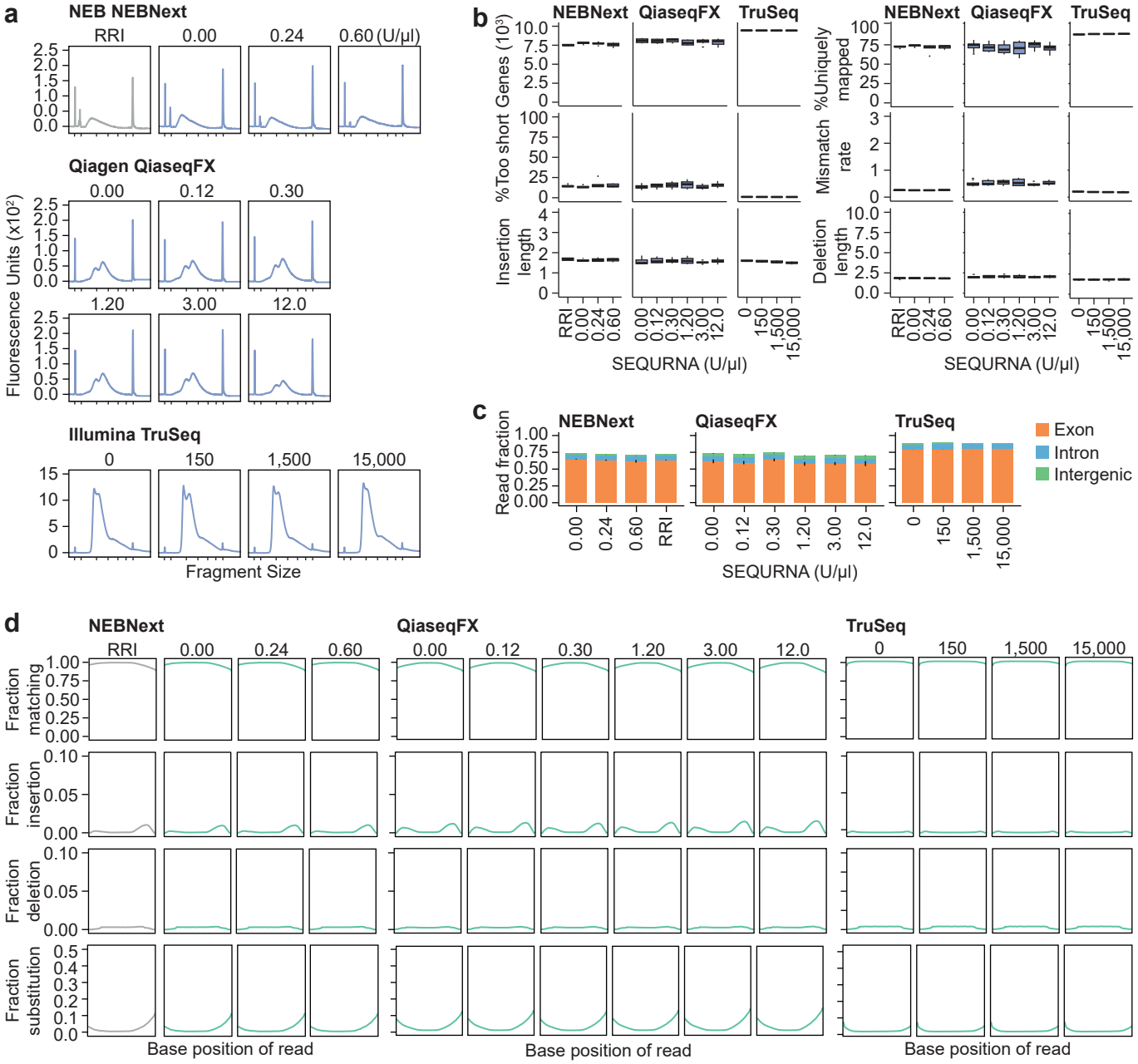


Supplementary Figure 6. Cell type marker gene expression in spleen and liver cells.
a. Expression-level heatmap of top 30 variable genes for each UMAP-generated cluster from single cell RNA sequencing of mouse liver (n = 349) and spleen-derived (n = 368) cells, sorted by gene expression with left bars indicating cell cluster based on transcriptome signature, source tissue, and RNase inhibitor condition (3 or 4.5 U/µl SEQURNA and standard SS2 with recombinant inhibitor), **b.** Mouse liver spleen-cell UMAP coloured by expression level of selected variable genes for each cell clusters: *Cd79a*, *Trbc1*, *Ahsg*, and *Cd300a*.



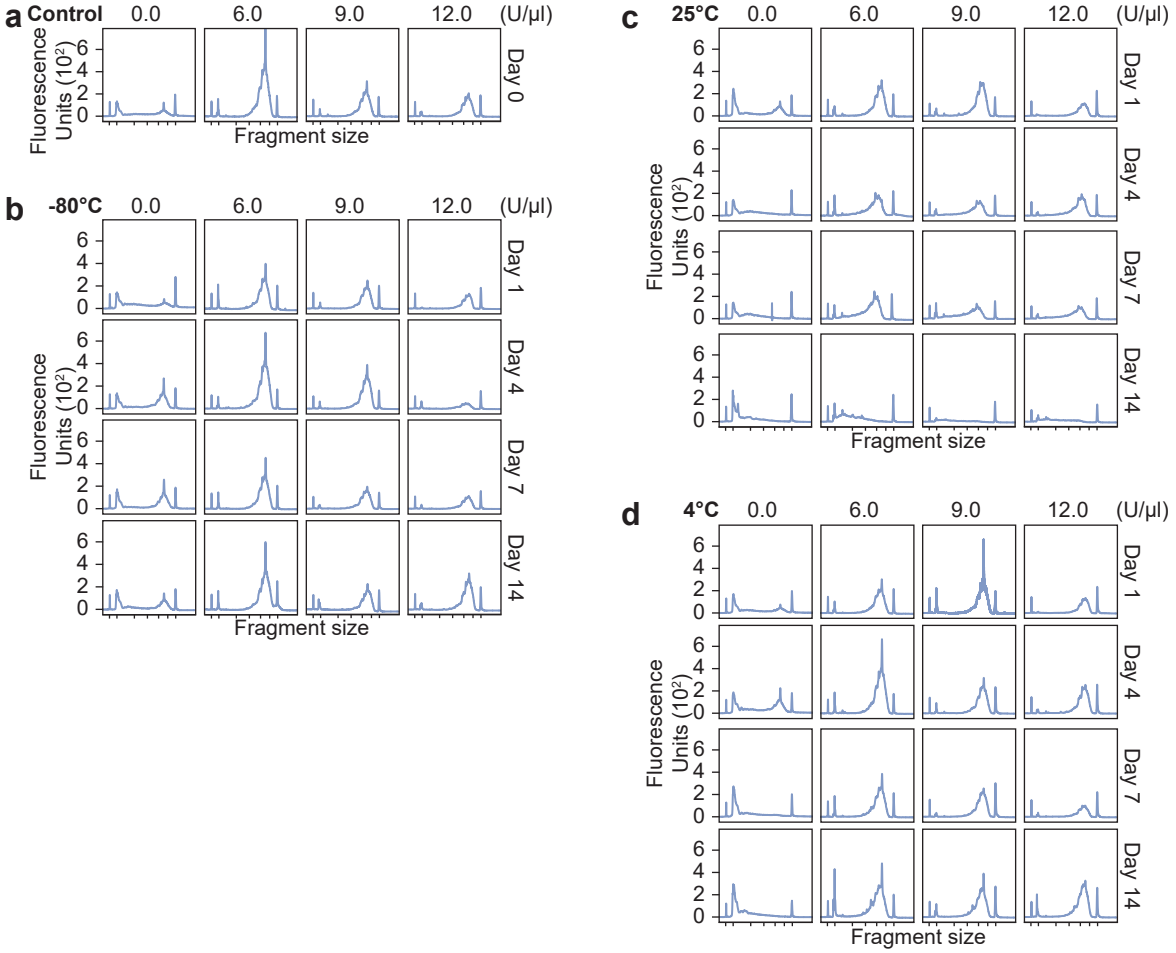
Supplementary 7. Mapping statistics of Smart-seq3 libraries generated using SEQURNA RNase inhibitor. **a.** Average percent (dot) and range (whiskers) of primer-dimer in Smart-seq3 (SS3) cDNA libraries from 100 pg of mouse total RNA and individual HEK293FT cells (integration range 20–50 bp). The inset corresponds to a zoom of the figure from 0–15%, highlighting the trend of decreasing primer-dimer occurrence with increasing concentration of SEQURNA up to the higher end of the "optimal range". **b.** Box plots of percent uniquely mapped to the reference genome (top), percent sequencing reads too short to be mapped (middle), and percent sequencing reads unmapped (not due to fragments too short or too many mismatches) (bottom) for 100 pg mouse total RNA and individual HEK293FT cells. **c.** Box plot of average insertion length (top), average deletion length (middle), and rate of mismatch per base (bottom) for 100 pg mouse total RNA and individual HEK293FT cells. **d.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the genome for 100 pg mouse total RNA and individual HEK293FT cells. **e.** Line plot of fraction of bases along sequencing reads with a deletion, with an insertion, matching the genome, and with a substitution indicating read quality for libraries from 100 pg mouse RNA and individual HEK293FT cells. **f.** Normalized gene body coverage of mapped reads along transcripts for libraries from 100 pg mouse total RNA and individual HEK293FT cells.

Supplementary Figure 8



Supplementary Figure 8. SEQURNA compatibility with commercial RNA-seq kits.
a. Bioanalyzer plots of final RNA sequencing libraries generated from the NEBNext RNA Sequencing Kit with 100 pg of mouse total RNA and varying amounts of SEQURNA (0, 0.24, or 0.6 U/µL) or RRI added in the lysis step. **b.** Bioanalyzer trances of sequencing libraries generated from the QIAseq FX Single Cell RNA Library Kit with 100 pg of mouse total RNA and varying amounts of SEQURNA (0, 0.12, 0.3, 1.2, 3, or 12 U/µl) added to the denaturation step. **c.** Bioanalyzer plots of RNA-sequencing libraries generated from the TruSeq RNA Sample Preparation Kit with 100 ng of mouse total RNA in a storage buffer containing varying amounts of SEQURNA (0, 150, 1,500, or 15,000 U/µl) prior to adding the RNA sample to the RNA purification oligo-dT beads. With this protocol, including a bead capture step of polyadenylated RNA, high SEQURNA concentrations were considered, evaluating whether detrimental carryover to the following library preparation protocol occurred. Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs. **d.** Box plot of number of genes detected for each inhibitor condition used to each kit. **e.** Box plot of percent uniquely mapped sequencing reads to the reference genome for each inhibitor condition for the NEBNext (left), QiaseqFX (center), and TruSeq (right) kits. **f.** Box plot of percent sequencing reads too short to be mapped to the reference genome for kit samples. **g.** Box plot of the mismatch rate length for kit samples. **h.** Box plot of the average insertion length for kit samples. **i.** Box plot of the average deletion length for kit samples. **j.** Stacked bar plot of fraction reads mapping to exonic, intronic, or intergenic regions of the mouse genome. **k.** Line plot of fraction of bases along sequencing reads matching the genome, indicating sequencing read quality. **l.** Line plot of fraction of bases along sequencing reads with insertion for kit samples. **m.** Line plot of fraction of bases along sequencing reads with deletion for kit samples. **n.** Line plot of fraction of bases along sequencing reads with substitution for kit samples.

Supplementary Figure 9



Supplementary Figure 9. SEQURNA delays RNA degradation of cells stored long-term in Smart-seq2 lysis buffer.

a. Bioanalyzer traces of control Smart-seq2 (SS2) cDNA libraries from HEK293FT cells FACS-sorted into lysis buffer containing either no inhibitor, 6, 9 or 12 U/µl of SEQURNA, where cDNA was generated immediately after cell sorting (day 0). **b.** Bioanalyzer traces of SS2 cDNA libraries from HEK293FT cells FACS-sorted into lysis buffer containing various concentrations of inhibitor processed into libraries after storage at -80°C for 1, 4, 7, or 14 days (top to bottom). **c.** Bioanalyzer traces of SS2 cDNA libraries from HEK293FT cells FACS-sorted into lysis buffer containing various concentrations of inhibitor after storage at 25°C for 1, 4, 7, and 14 days (top to bottom). **d.** Bioanalyzer traces of SS2 cDNA libraries from HEK293FT cells FACS-sorted into lysis buffer containing various concentrations of inhibitor after storage at 4°C for 1, 4, 7, and 14 days (top to bottom). Tick marks on the x-axis correspond to 35, 100, 300, 500, 1000, 3000, and 10380 base pairs.