

ARTICLE **OPEN**

Cost and time-efficient construction of a 3'-end mRNA library from unpurified bulk RNA in a single tube

Jungwon Choi^{1,4}, Jungheun Hyun^{1,4}, Jieun Hyun^{1,4}, Jae-Hee Kim¹, Ji Hyun Lee^{2,3}✉ and Duhee Bang¹✉

© The Author(s) 2024

The major drawbacks of RNA sequencing (RNA-seq), a remarkably accurate transcriptome profiling method, is its high cost and poor scalability. Here, we report a highly scalable and cost-effective method for transcriptomics profiling called Bulk transcriptOme profiling of cell Lysate in a single pot (BOLT-seq), which is performed using unpurified bulk 3'-end mRNA in crude cell lysates. During BOLT-seq, RNA/DNA hybrids are directly subjected to tagmentation, and second-strand cDNA synthesis and RNA purification are omitted, allowing libraries to be constructed in 2 h of hands-on time. BOLT-seq was successfully used to cluster small molecule drugs based on their mechanisms of action and intended targets. BOLT-seq competes effectively with alternative library construction and transcriptome profiling methods.

Experimental & Molecular Medicine (2024) 56:453–460; <https://doi.org/10.1038/s12276-024-01164-8>

INTRODUCTION

RNA-seq is used routinely in biomedical research and can be readily performed using commercially available kits^{1,2}. Although the cost of DNA sequencing has substantially decreased over time, the cost and complexity of standard library preparation remain significant concerns^{3–5}. The main steps performed to construct libraries from bulk transcriptome profiling through RNA-seq include RNA extraction, reverse transcription, second-strand complementary DNA (cDNA) synthesis, tagmentation, and PCR amplification. Typically, each of these sequential reaction steps is performed using a commercial kit or a set of specific reagents and must be properly completed before the next step is performed^{6–8}. After each experimental step, intermediate products must be purified and verified. Even though the cost of sequencing has decreased, the cumulative labor, time, and cost needed to construct NGS libraries using commercially available kits can be prohibitive many libraries are needed. Because the costs of preparing NGS libraries and sequencing increase with increasing sample numbers, financial considerations often limit the choice of experimental design.

Through whole RNA-seq-based construction of NGS libraries, transcriptomes can be thoroughly analyzed. However, these protocols are usually associated with longer preparation times and increased costs^{9,10}. On the other hand, 3'-end mRNA-seq, which limits DNA sequence analysis to the 3'-terminal region of mRNA transcripts, is an attractive approach for increasing throughput¹¹. Furthermore, more samples can be combined per NGS run, and more relative read depth per gene can be achieved with 3'-end mRNA-seq than with whole transcriptome sequencing. To exploit these advantages, several 3'-end mRNA-seq library preparation protocols as well as analytical procedures have been developed.

Several newly developed 3'-end mRNA-seq protocols require smaller amounts of reagents and fewer commercial kits, resulting in reduced costs. For example, Bulk RNA Barcoding and sequencing (BRB)-seq is a rapid, inexpensive method for generating 3'-end mRNA-seq libraries from bulk RNA¹². BRB-seq uses “in-house”-produced Tn5 transposase for tagmentation of double-stranded cDNA, resulting in a substantial reduction in cost. However, in the BRB-seq protocol, RNA must be purified using a commercially available kit. Then, purified RNA is used for reverse transcription for the next reaction step. On the other hand, Digital RNA with perturbation of Genes (DRUG)-seq is a unique 3'-end mRNA-seq method that uses bulk RNA in cells that are directly lysed in mild lysis buffer in a 96-well plate format¹³. Bulk RNA in the lysate is then individually reverse-transcribed and pooled prior to purification. This process minimizes hands-on time and work at the start of the procedure; thus, multiple samples can be processed simultaneously. However, DRUG-seq requires sample pre-amplification and multiple purification steps, making it a slower and more labor intensive method (Supplementary Fig. 1).

Here, we present Bulk transcriptOme profiling of cell Lysate in a single pot (BOLT-seq), a new 3' mRNA-seq method that incorporates some features of BRB-seq and DRUG-seq. BOLT-seq is a highly scalable cost-effective transcriptome sequencing method targeting 3'-end sequences of unpurified bulk mRNA directly in the lysate of up to 1000 cells; BOLT-seq also uses reverse transcriptase (RT)¹⁴ and Tn5 transposase¹⁵ purified in our laboratory (in-house) as well as in-house prepared reaction buffers to reduce experimental cost. In the BOLT-seq protocol, purification of intermediate products is not necessary until the library amplification step; therefore, the experiment be carried out in a single tube. By subjecting RNA/DNA hybrid duplexes to Tn5 transposase-mediated tagmentation¹⁶, second-strand cDNA

¹Department of Chemistry, Yonsei University, Seoul, Republic of Korea. ²Department of Clinical Pharmacology and Therapeutics, College of Medicine, Kyung Hee University, Seoul, Republic of Korea. ³Department of Biomedical Science and Technology, Kyung Hee University, Seoul, Republic of Korea. ⁴These authors contributed equally: Jungwon Choi, Jungheun Hyun, Jieun Hyun. ✉email: hyunihyuni@khu.ac.kr; duheebang@yonsei.ac.kr

Received: 7 June 2023 Revised: 19 October 2023 Accepted: 12 November 2023

Published online: 27 February 2024

synthesis and pre-amplification can be omitted from the BOLT-seq protocol. BOLT-seq requires minimal hands-on labor, and a 3'-end mRNA library can be constructed in less than one day for under US \$1.40 per sample (excluding the cost of NGS sequencing) [Supplementary Table 1] while fulfilling standard expectations for the quality of DNA sequence and gene expression data in multiple cell lines. To demonstrate the potential of the method, BOLT-seq was used to analyze the effects of 35 small molecule drugs on the transcriptome of cells in a 96-well microtiter plate. The data from this experiment were successfully used to cluster the drugs by their mechanisms of action and intended targets.

MATERIALS AND METHODS

Cell lines

We purchased the HEK293T, A549, and NIC-H358 cell lines from the Korean Cell Line Bank (KCLB No. 21573, KCLB No. 10185, KCLB No. 25807). HEK293 cells were maintained in DMEM or RPMI Medium 1640 (Gibco, USA) with 10% fetal bovine serum (Gibco, USA) and 1% penicillin/streptomycin (Gibco, USA). Cells were cultured at 37 °C under 5% CO₂.

Purification of reverse transcriptase and Tn5 transposase

The reverse transcriptase used in this study was a variant of M-MuLV RT cloned and inserted into the pET-28a(+) vector (Millipore Sigma, USA), which carries a kanamycin-resistance gene. M-MuLV was expressed and purified according to the protocol published by Baranauskas et al.¹². The Tn5 transposase used in this study was from pTXB1-Tn5, which expresses Tn5 transposases and an ampicillin resistance gene [gift from Rickard Sandberg (Addgene plasmid #60240; <http://n2t.net/addgene:60240>; RRID: Addgene_60240)]. Tn5 transposase was expressed and purified according to the protocol published by Picelli et al.¹⁵.

RNA-seq library preparation

Cells were cultured in 96-well microtiter plates, and total RNA was purified using an RNeasy mini kit (Qiagen, Netherlands). Total RNA quality was analyzed using the 4150 TapeStation System (Agilent, USA). NEBNext Ultra II RNA libraries were prepared from 200 ng purified total RNA according to the manufacturer's instructions (NEB, USA). Libraries were sequenced by NovaSeq 6000.

BOLT-seq library preparation

To prepare BOLT-seq libraries, 96-well plates were seeded with 10,000 cells, incubated overnight, washed twice with DPBS (Gibco, USA), and lysed in 60 µL of lysis buffer containing 0.3% IGEPAL CA-630 (Sigma Aldrich, USA). The plates were incubated for 30 min with shaking at 800 RPM. Cell lysate (6 µL) was transferred into a PCR tube for library preparation. RNA was denatured by adding 1 µL RT-Mix-A containing 10 µM anchored oligo(dT) 30-P7 RT primer (Integrated DNA Technologies, South Korea) and 1 µL 1 mM dNTP mix (Thermo Fisher, USA) to the cell lysate. Then, the mixture was incubated at 65 °C for 5 min and quickly cooled on ice for 3 min. The RT reaction was performed by adding 7 µL RT-Mix-B containing 117 mM Tris-HCl (pH 8.3), 175 mM KCl, 7 mM MgCl₂, 23 mM DTT, 23% PEG8000 (w/v) (bioPLUS, USA), 5 U RNase OUT Ribonuclease inhibitor (Thermo Fisher, USA), 0.5 µL in-house purified M-MuLV RT and incubation at 50 °C for 60 min. The reaction was inactivated by incubation at 80 °C for 10 min. No purification step was needed. DNA/RNA hybrid duplexes were subjected to tagmentation by adding 5 µL of TD-Mix containing 40 mM Tris-HCl (pH 7.5), 20 mM MgCl₂, 30% PEG8000 (w/v) (bioPLUS, USA), 20% tetraethylene glycol (Thermo Scientific, USA), and 0.5 µL of in-house purified Tn5 transposase and incubated at 55 °C for 30 min. Tagmentation was stopped by adding 5 µL of 0.2% SDS. No purification step was needed. Gap-filling and PCR amplification were performed by adding 25 µL PCR-Mix containing 5x HiFi Fidelity Buffer (Kapa Biosystems, UK), KAPA dNTP Mix (Kapa Biosystems, UK), KAPA HiFi HotStart DNA Polymerase (Kapa Biosystems, UK), 0.5 µL in-house purified RT, and 2 µL NGS indexed primers. Gap-filling and PCR were performed as follows: 50 °C for 10 min for gap-filling, 95 °C for 3 min for initial denaturation and inactivation of RT, 18 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, 72 °C for 3 min for final extension. The indexed products were purified at 0.6X with SpeedBead Magnetic Carboxylate Modified Particles (GE Healthcare, UK). Library products were quantified with Qubit 2.0, and library length distribution was analyzed with the 4150 TapeStation System (Agilent, USA).

RNA-seq library data analysis

Raw NGS reads were trimmed to remove low-quality bases and adapter sequences using Trimmomatic 0.39¹⁷ (ILLUMINACLIP:1:30:7, LEADING:3, TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:20). Subsequent processes, including trimming, paired-end, and single-end options, were applied according to RNA-seq methods. Trimmed reads were mapped to the human genome (GRCh37/hg19) using STAR 2.7.6a¹⁸ (--outSAMmapqUnique 60). Only uniquely mapped aligned reads were selected and used for subsequent analyses. bam files were subsampled identically in each experiment. The subsampling size is specified in the figure legend as needed. SAMtools 1.11¹⁹ was used to sort uniquely mapped reads and for subsampling. Unique subsampled mapped reads were assigned to the set of 42149 annotated genes in the GRCh37 assembly of the human genome using featureCount 1.6.5²⁰ from the package Subread. Raw read counts from HEK293T and A549 cells were analyzed with DESeq2²¹ to detect DE genes.

Gene set expression analysis

To test the GSEA of the H358 drug reaction performed with the NEBNext Ultra II RNA kit and BOLT-seq, a KRAS dependency gene-related gene set of previously published data^{22–26} was obtained. Enrichment of gene sets was analyzed using GSEA v3.0²⁷ and the MSigDB gene set database from published data²⁸ for the Hv6.0 collection of hallmark sets, the C2v6.0 collection of chemical and genetic perturbations (CGP) and canonical pathways (CP), the C5v6.0 collection of GO biological and molecular functions, and the C6v6.0 collection of oncogenic signatures.

RESULTS

Bulk transcriptome profiling of cell lysate in a single pot—BOLT-seq library preparation and workflow

As a method for gene expression profiling, BOLT-seq is a streamlined, inexpensive, and time-efficient; utilizes a 96-well microtiter plate format; does not require intermediate purification steps; and does not involve commercially available kits. The workflow and sequential steps of BOLT-seq library preparation are shown schematically in Fig. 1. To summarize, after cells were lysed in the well of a 96-well plate, reverse transcription was performed using in-house purified Moloney murine leukemia virus (M-MuLV) reverse transcriptase and bead-anchored oligo-dT primers [Supplementary Table 2]. Note that a substantial cost reduction was achieved by using in-house purified M-MuLV and in-house prepared reaction buffers. To ensure optimal RT efficiency, the activity of in-house purified M-MuLV RT was tested with different reaction buffers¹⁴. After the RT step, RNA/DNA hybrid duplexes were used for the Tn5 transposase-mediated tagmentation reaction; thus, second-strand cDNA synthesis could be omitted¹⁶. As the BOLT-seq protocol eliminates the step in which RNA transcripts are converted to double-strand DNA, the experimental cost and the time needed for library preparation are also reduced. Further cost reduction was achieved by purifying Tn5 transposase and preparing its reaction buffer in-house according to a protocol published by Picelli et al.¹⁵. The products of Tn5 transposase-mediated tagmentation were then used without purification in the subsequent gap-filling and PCR amplification steps. Thus, BOLT-seq is a streamlined procedure that reduces the total time for NGS library construction to 4 h at a dramatically reduced cost to US \$1.40 per well.

Performance of BOLT-seq

To readily compared the method to traditional RNA-seq methods, BOLT-seq was performed under several different experimental conditions. Initially, the variables tested included the number of input cells per well and the presence or absence of a crowding agent, such as polyethylene glycol 8000 (PEG8000). HEK293T and A549 cell lines were used to compare each experimental condition. To test the robustness of BOLT-seq, initial experiments were performed using an appropriate number of cells per well, and in-house purified Tn5 transposase was used for RNA/DNA hybrid tagmentation. Finally, the index PCR cycle was fixed at 18 cycles for comparison.

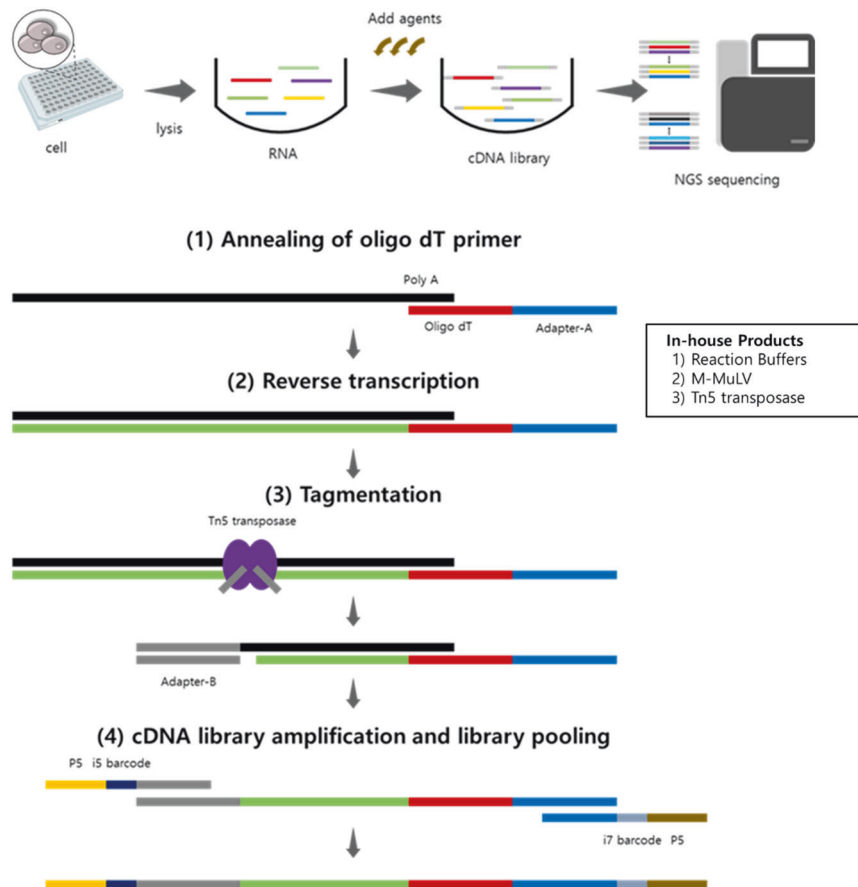


Fig. 1 Overall BOLT-seq scheme. Bulk transcriptome profiling of cell lysate in a single poT—BOLT-seq. The steps of BOLT-seq are shown schematically. Sequential reactions were carried out in a single well of a 96-well microtiter plate with no intermediate purification steps. After the final library amplification step, products are purified from each well and pooled as desired.

As reported, the efficiency of the previous protocol decreased when >1000 cells were lysed in a single well²⁹. Therefore, the BOLT-seq libraries prepared in this study were derived from no more than 1000 cells per well, and the performance of BOLT-seq using 100, 500, and 1000 cells per well was compared. For libraries derived from 100 cells, many NGS sequencing reads had to be discarded due to unacceptably short length (Supplementary Fig. 2a); as a result, no additional experiments were performed using 100 cells per well. More genes were detected in libraries prepared from 1000 cells than in libraries prepared from 500 cells, but this difference was not significant (Supplementary Fig. 2b). For consistency and comparability of reaction controls, all subsequent experiments in this study used 1000 cells per well.

PEG8000 is frequently added to RT reactions to increase molecular crowding, but it has also been used to effect conformational change and facilitate Tn5 transposase-mediated tagmentation of RNA/DNA hybrid duplexes¹⁶. In this study, no significant difference in gene detection was observed when BOLT-seq was performed in the presence of 0% or 9% PEG8000 (Supplementary Fig. 2c). However, a significantly higher DNA yield was obtained during the final PCR stage of BOLT-seq in the presence of 9% PEG8000. Therefore, all subsequent BOLT-seq reactions were performed in the presence of 9% PEG8000.

The reproducibility of BOLT-seq was evaluated by comparing the number of expressed genes detected in common between three independent replicate experiments in two different cell lines. The results showed that 11,150 genes were detected in all three replicates in HEK293T cells, and 10,823 genes were detected in all three replicates in A549 cells, representing 56.5% and 56.4% of all genes detected in HEK293T and A549 cells, respectively

(Fig. 2a). In addition, normalized gene read counts were compared pairwise for all three replicate experiments in HEK293T and A549 cells, and correlation values ranged from 0.980 to 0.994 in HEK293T cells and from 0.988 to 0.993 in A549 cells (Fig. 2b).

BOLT-seq data were also verified using variable amounts of a spike-in reference standard and methods developed by the US National Institute of Standards and Technology-sponsored External RNA Controls Consortium (ERCC). For each experimental replicate in HEK293T and A549 cells, ERCC read counts were compared with expected ERCC values, and BOLT-seq data were compared with data from the NEBNext Ultra II RNA Kit. The results showed that normalized ERCC counts between replicates were highly correlated ($r > 0.999$) (Fig. 2c, Supplementary Fig. 3a) for the BOLT-seq and NEBNext methods, indicating that both methods achieve a high level of reproducibility. The observed ERCC counts and expected ERCC amounts were also highly correlated (Fig. 2d, Supplementary Fig. 3b), with correlation values ranging from 0.942 to 0.976 for NEBNext and from 0.924 to 0.950 for BOLT-seq. These results confirm that BOLT-seq achieves an acceptable and desirable level of performance. Transcript detection efficiency was also evaluated by plotting the ERCC probe concentration and length using the RNA-seq method; the results indicate that selective sequencing does not occur as a function of probe length. The results also show dropout, which is depicted as black circles in Fig. 2e, at low ERCC probe concentrations independent of the ERCC probe length. For replicates with BOLT-seq, the ERCC probe mix was spiked at ratios of 1:30 or 1:100 to demonstrate that low-concentration ERCC probes could be better detected and sequenced in samples prepared at a ratio of 1:30 than 1:100. These results indicate that variations in probe length or

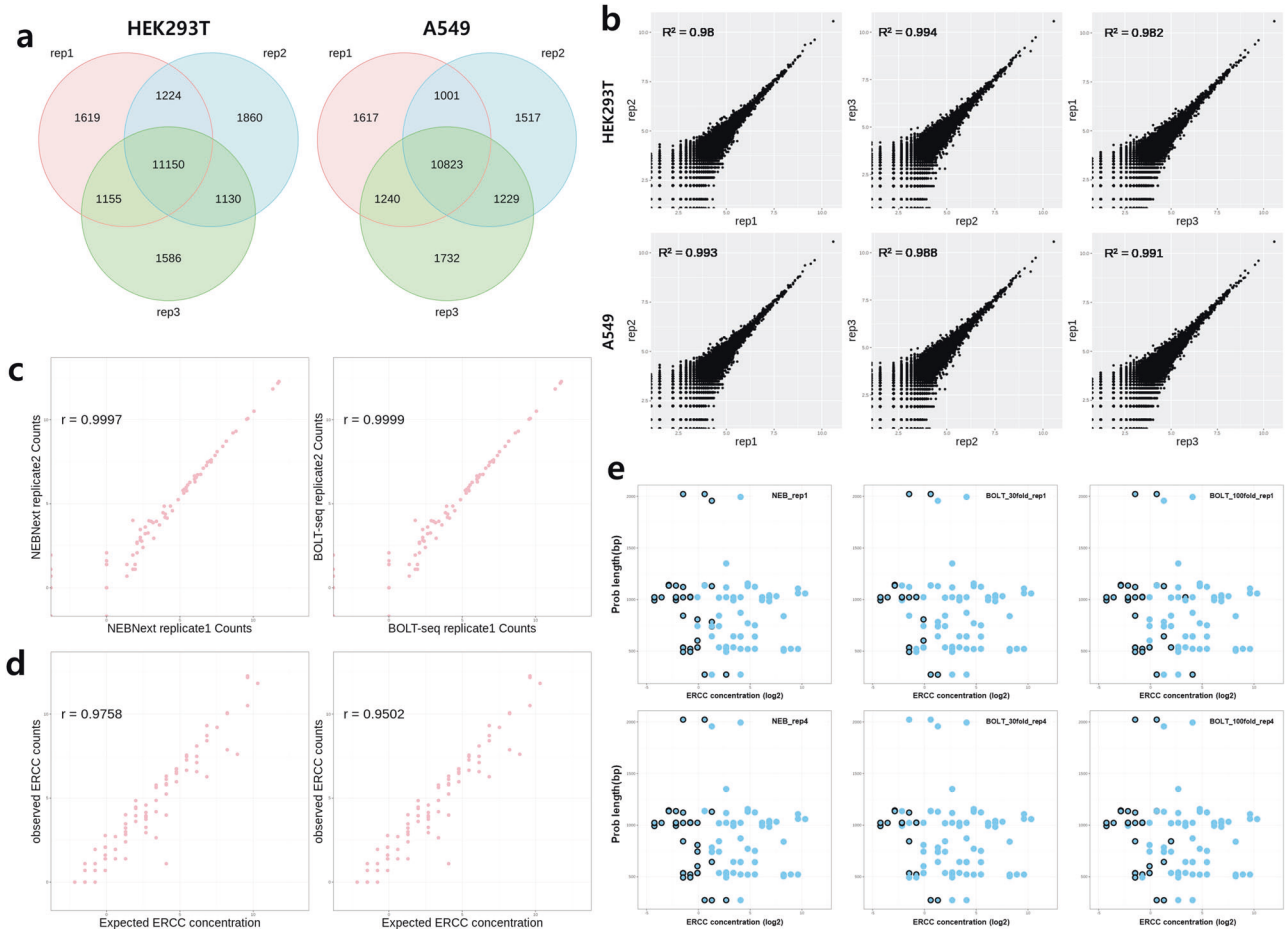


Fig. 2 Performance of BOLT-seq. Reproducibility of the BOLT-seq method. All experimental replicates were subsampled to 1 M. Venn diagrams in (a) represent overlapping subsets of genes detected in each of three replicate BOLT-seq samples using 1000 HEK293T (left) or A549 (right) cells per well in a 96-well microtiter plate. **b** Normalized gene read counts from each of the three replicate experiments per cell line shown in (a) were compared pairwise. Correlation values (r) for each pairwise comparison are shown in the upper left corner of each panel. Upper panels, HEK293T cells; lower panels, A549 cells. **c** Correlation between unique normalized ERCC read counts in two experimental BOLT-seq replicates using 1000 HEK293T cells. Left panel, NEBNext; right panel, BOLT-seq. **d** Correlation between observed and expected ERCC read counts using NEBNext (left) or BOLT-seq (right). **e** The relationship between the ERCC concentration and probe length using NEBNext (left), BOLT-seq 1/30 (middle), and BOLT-seq 1/100 (right). Undetected dropouts are indicated by a black circle.

concentration are not associated with sequencing bias during BOLT-seq. Thus, compared to traditional mRNA-seq methods, BOLT-seq is simpler, less labor intensive, and less expensive.

Characterization of M-MuLV reverse transcriptase purified in-house

BOLT-seq is performed with in-house purified M-MuLV RT instead of commercially available RT to lower the cost of library preparation. The quality of in-house purified M-MuLV RT was established by comparing the number of DE genes detected by BOLT-seq with in-house M-MuLV RT, commercially available Maxima™-H RT, or SuperScript™ IV (SSIV) RT. These data were also compared with data generated with NEBNext. Venn diagrams representing the number of DE genes detected in each condition are shown in Supplementary Fig. 4a. The results show that in-house M-MuLV RT detects fewer DEs than SSIV RT, but it also detects more DEs than Maxima™-H RT. Therefore, while the performance of in-house M-MuLV RT is not optimal, it performs comparably to commercially available preparations of M-MuLV RT. We also compared the log₂-fold change in DE genes with BOLT-seq or NEBNext (Supplementary Fig. 4b), which revealed that the results with BOLT-seq and NEBNext ($r = 0.961$) and the log₂-fold change (lfc) observed with BOLT-seq or NEBNext using in-house M-MuLV RT are also highly correlated ($r = 0.944$). These results

suggest that the function and activity of in-house M-MuLV RT are comparable to the function and activity of commercially available RT preparations. Thus, in-house M-MuLV RT can be safely used without compromising data quality while considerably reducing the cost of library construction with BOLT-seq.

Comparing BOLT-seq to other RNA-seq methods

Next, the performances of BOLT-seq, TRACE-seq³⁰, and NEBNext were compared. For these experiments, 200 ng of total RNA was purified from HEK293T and A549 cells and used for library construction with TRACE-seq or NEBNext, whereas BOLT-seq was performed using 1000 HEK293T or A549 cells per well. For each experimental method, DE genes were identified by comparing HEK293T and A549 cells using DESeq2. A p value less than 5×10^{-6} and $|\log_2\text{-fold change}| > 1$ were used as cutoff criteria, and the results are represented in the Venn diagrams in Fig. 3a. TRACE-seq and NEBNext detected 2288 and 3296 DE genes, respectively, while 1007 DE genes were detected by both NEBNext and BOLT-seq, representing 85% of all DE genes detected by BOLT-seq (Fig. 3a, left). Similarly, 997 DE genes were detected by both TRACE-seq and BOLT-seq, representing 84.1% of all DE genes detected by BOLT-seq (Fig. 3a, right). Although BOLT-seq detects far fewer DE genes than canonical RNA-seq methods, >80% of the DE genes detected by BOLT-seq are also detected by TRACE-seq

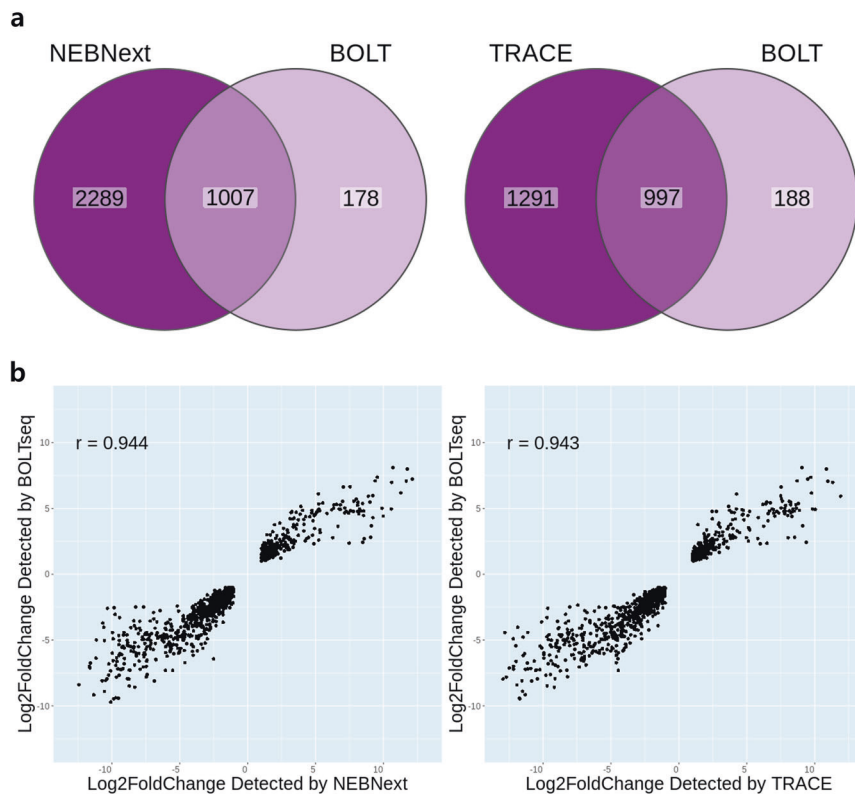


Fig. 3 Comparison between BOLT-seq and the canonical RNA-seq method. Comparison between BOLT-seq and the canonical TRACE-seq and NEBNext methods. All replicates were subsampled to 1 M (replicates: NEBNext $n = 3$, TRACE-seq $n = 3$, BOLT-seq $n = 3$). **a** Venn diagrams show overlapping subsets of DE genes detected with NEBNext and BOLT-seq (left) or TRACE-seq and BOLT-seq (right). NEBNext and TRACE-seq were performed using 200 ng of total RNA purified from HEK293T or A549 cells. BOLT-seq was performed using the lysates of 1000 HEK293T or A549 cells. **b** Correlation between the log₂-fold change (lfc) of DE genes detected by NEBNext and BOLT-seq (left) or TRACE-seq and BOLT-seq (right). Cutoff criteria were threshold: $|lfc| > 1$, $p\text{-adj} < 0.05$.

and NEBNext. In addition, the lfc values of DE genes detected by NEBNext or TRACE-seq and BOLT-seq were highly correlated ($r > 0.94$ for both; Fig. 3b, left and right). Therefore, the reproducibility and performance of BOLT-seq and established RNA-seq methods are comparable.

Application of BOLT-seq

By continuously adding reagents into a single tube without purifying the reaction products between sequential reaction steps, the cost and time needed for large-scale preparation of NGS libraries can be reduced. Here, a proof-of-concept experiment was performed to show that BOLT-seq can be used in a large-scale screen for drug-induced perturbation of gene expression in NIC-H358 cells, which are KRAS G12C mutant non-small cell lung carcinoma cells. The cells were exposed to 35 drugs [Supplementary Table 3] in triplicate on two experimental days with 9 replicates of the DMSO control, generating 213 data points per sample (Fig. 4a). The NGS library was prepared through the BOLT-seq method, and then gene expression profiling was performed according to the bioinformatic pipelines from the DRUG-seq method^{13,31,32}, followed by unsupervised clustering to identify genes with similar drug-induced perturbations^{33,34} (Fig. 4b). Clusters appeared to reflect the type of drug (Supplementary Fig. 5a) but were not influenced by the date of drug treatment (Supplementary Fig. 5b). For example, Cluster 1 included single drugs or drug combinations that target the mitogen-activated protein kinase (MAPK) pathway, which includes the KRAS and MEK proteins. Drugs that inhibit KRAS or MEK inhibit growth and induce apoptosis in NIC-H358 cells^{33–35}. AMG510 is a drug that targets KRAS³⁶, while trametinib targets MEK³⁴, both of which are genes in the MAPK pathway. As AMG510 and trametinib are

involved in the same pathway, they are expected to cluster together. To confirm this, we used the DEseq normalization method²¹ and analyzed the normalized and p-adjusted values of gene expression in AMG510- and Trametinib-perturbed samples. Then, the top 20 significant genes for these drugs were identified based on the lowest p-adjusted values (Fig. 4c). BOLT-seq data were also used to identify genes in the ERK pathway^{37,38}, such as FOSL1 and CCND1, in cells treated with AMG510 and trametinib. Gene set enrichment analysis (GSEA)²⁷ was performed to analyze the perturbation of gene expression in NIC-H358 cells treated with AMG510 or trametinib. The gene set (see Methods) of the results analyzed by GSEA, which is the result of gene expression of NIC-H358 cells treated with ARS1620 and trametinib, was used. GSEA confirmed that AMB510 and trametinib downregulate the expression of KRAS-related genes, including E2F transcription factors, the MYC regulatory network, ERK activation, and KRAS dependency signatures^{22–26} (Fig. 4d).

In summary, BOLT-seq is a novel RNA-seq method that facilitates large-scale transcriptome profiling and drug screening experiments by dramatically reducing the labor, time, and cost of library construction relative to canonical methods. The present study also demonstrates that BOLT-seq performs as well as canonical RNA-seq methods.

DISCUSSION

Preparing NGS libraries using commercially available kits is labor intensive and very expensive; with the NEBNext Ultra II RNA kit, for example, each sample costs up to US\$45–\$47. In comparison, one person can generate libraries for up to 96 samples at a time with BOLT-seq, which is complete in 4 h with 2 h of hands-on time and

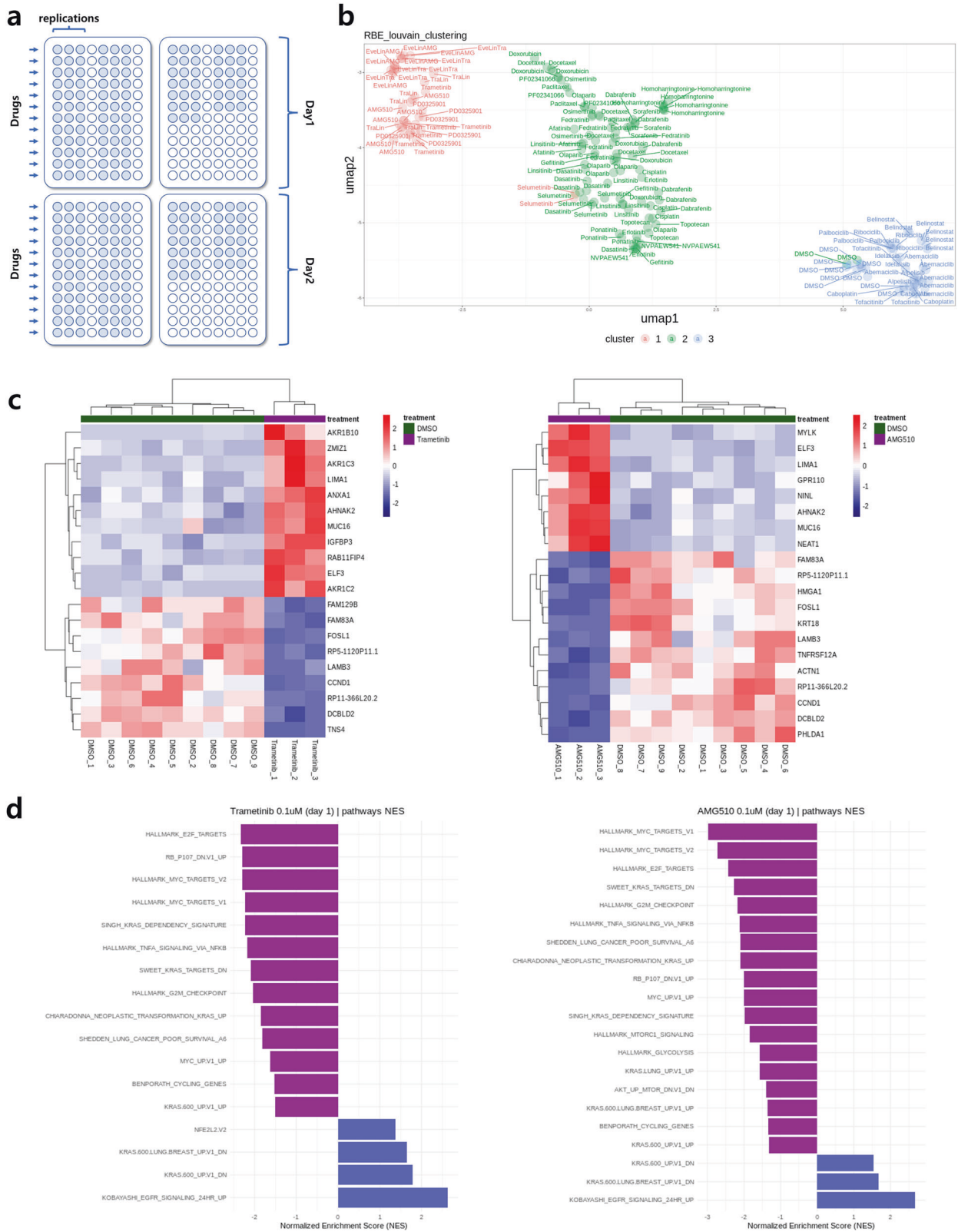


Fig. 4 Application of BOLT-seq. a Schematic diagram of drug screening with BOLT-seq. **b** Uniform manifold approximation and projection (UMAP) clustering of data for 35 drugs x 2 days x three replicates (DMSO $n = 9$). **c** Top 20 significant DE genes detected in cells exposed to AMG510 or Trametininb. **d** Gene Set Enrichment Analysis for AMG510 and Trametininb.

costs US \$1.40 per sample. Because BOLT-seq libraries are prepared from unpurified cell lysates using Tn5 transposase-mediated tagmentation of RNA/DNA hybrid duplexes while excluding intermediate purification steps, the time needed for library preparation is dramatically reduced. Furthermore, as M-MuLV RT and Tn5 transposase and their respective buffers are purified in-house, the cost of library preparation is dramatically reduced. In-house-produced M-MuLV RT and Tn5 transposases can be replaced with commercial brand enzymes. However, the final cost of BOLT-seq using commercial brand enzymes is higher than the original BOLT-seq cost, as shown in Supplementary Table 4. As a result, BOLT-seq facilitates large-scale preparation of NGS libraries from unpurified 3'-end mRNA at minimal cost within 4 h; thus, BOLT-seq is a cost- and time-efficient method for large-scale 3'-end mRNA-seq studies.

Comparing the performance of BOLT-seq, TRACE-seq, and NEBNext is challenging. NEBNext is widely used and is the gold standard full-transcriptome mRNA-seq method. Because BOLT-seq is a 3'-end mRNA-seq method, the correlation coefficient for comparing DE genes with BOLT-seq and NEBNext methods ($r = 0.94$) is slightly lower than when TRACE-seq and NEBNext are compared ($r = 0.96$) because TRACE-seq and NEBNext are full-transcriptome mRNA-seq methods. Nevertheless, a correlation coefficient of 0.94 between BOLT-seq and NEBNext is high, given the difference in the methodology.

The performance of BOLT-seq in large-scale drug screening (Fig. 4) is promising. Figure 4 shows successful clustering of drugs by mechanism of action and intended target based on BOLT-seq data. For example, drugs that inhibit MEK and KRAS and downregulate the MAPK pathway inhibit the growth of KRAS-mutant NCI-H358 cells. When a similar screen was performed with NEBNext, more potential drug targets were detected than with BOLT-seq. However, BOLT-seq identified small molecule drug-induced changes in gene expression as well as NEBNext. Due to the relatively lower cost and reduced time needed to prepare and screen libraries with BOLT-seq, 35 different small molecule drugs were screened in triplicate on 2 experimental days. However, the cost of performing the same experiment with NEBNext would likely be prohibitive.

We note that the BOLT-seq method has several limitations and could be improved considerably. First, since BOLT-seq does not cover the whole transcriptome, it is not suitable for studies that focus on alternative splicing events, long noncoding RNAs, or single nucleotide variants in the 5' end. Nevertheless, reading only the 3' end plays greatly helps save time and money and provides sufficient information for accurate gene expression measurements. Second, although BOLT-seq costs less than traditional methods, the cost should likely decrease further as the cost of NGS sequencing continues to decrease. Additional cost reduction could be achieved by replacing commercially available DNA polymerase with in-house purified DNA polymerase. Finally, the methodological aspects of BOLT-seq can be improved. Currently, a high percentage of data are discarded when the raw NGS data are preprocessed. Optimizing the BOLT-seq reagents would increase the quality of the cDNA library. Therefore, we hope that the BOLT-seq method will be further improved through future investigation and development.

Much evidence suggests that NGS is an important and effective tool for biomedical research. Because BOLT-seq is a highly scalable, time- and cost-effective, plate-based 3'-end mRNA-seq method, it is expected to support many new and interesting multiscale transcriptomic experiments going forward.

REFERENCES

- Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
- Pickrell, J. K. et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
- Podnar, J., Deiderick, H., Huerta, G. & Hunicke-Smith, S. Next-generation sequencing RNA-Seq library construction. *Curr. Protoc. Mol. Biol.* **106**, 4.21.1–4.21.19 (2014).
- Song, Y. et al. A comparative analysis of library prep approaches for sequencing low input transcriptome samples. *BMC Genom.* **19**, 696 (2018).
- Masters, T. L. et al. Comparative evaluation of cDNA library construction approaches for RNA-Seq analysis from low RNA-content human specimens. *J. Microbiol. Methods* **154**, 55–62 (2018).
- Gubler, U. & Hoffman, B. J. A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269 (1983).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
- Cui, P. et al. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265 (2010).
- Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M. C. & Vitale, L. GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database* **2016**, baw153 (2016).
- Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Wilkening, S. et al. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* **41**, e65–e65 (2013).
- Alpern, D. et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.* **20**, 71 (2019).
- Ye, C. et al. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.* **9**, 4307 (2018).
- Baranauskas, A. et al. Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel.* **25**, 657–668 (2012).
- Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
- Di, L. et al. RNA sequencing by direct tagmentation of RNA/DNA hybrids. *Proc. Natl Acad. Sci. USA* **117**, 2886–2893 (2020).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
- Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
- Bild, A. H. et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2005).
- Dry, J. R. et al. Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res.* **70**, 2264–2273 (2010).
- Singh, A. et al. A gene expression signature associated with “K-ras addiction” reveals regulators of EMT and tumor cell survival. *Cancer Cell* **15**, 489–500 (2009).
- Sweet-Cordero, A. et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat. Genet.* **37**, 48–55 (2004).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Janes, M. R. et al. Targeting KRAS mutant cancers with a covalent G12C-specific inhibitor. *Cell* **172**, 578–589.e17 (2018).
- Le, A. V., Huang, D., Blick, T., Thompson, E. W. & Dobrovic, A. An optimised direct lysis method for gene expression studies on low cell numbers. *Sci. Rep.* **5**, 12859 (2015).
- Lu, B. et al. Transposase-assisted tagmentation of RNA/DNA hybrid duplexes. *eLife* **9**, e54919 (2020).
- Li, J. et al. DRUG-seq provides unbiased biological activity readouts for neuroscience drug discovery. *ACS Chem. Biol.* **17**, 1401–1414 (2022).
- McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. <https://doi.org/10.48550/ARXIV.1802.03426> (2018).
- Molina-Arcas, M. et al. Development of combination therapies to maximize the impact of KRAS-G12C inhibitors in lung cancer. *Sci. Transl. Med.* **11**, eaaw7999 (2019).

34. Gilmartin, A. G. et al. GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. *Clin. Cancer Res.* **17**, 989–1000 (2011).
35. Molina-Arcas, M., Hancock, D. C., Sheridan, C., Kumar, M. S. & Downward, J. Coordinate direct input of both KRAS and IGF1 receptor to activation of PI3 kinase in KRAS-mutant lung. *Cancer Cancer Discov.* **3**, 548–563 (2013).
36. Hong, D. S. et al. KRASG12C inhibition with sotorasib in advanced solid tumors. *N. Engl. J. Med.* **383**, 1207–1217 (2020).
37. Vallejo, A. et al. An integrative approach unveils FOSL1 as an oncogene vulnerability in KRAS-driven lung and pancreatic cancer. *Nat. Commun.* **8**, 14294 (2017).
38. Kwong, L. N. et al. Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. *J. Clin. Invest.* **125**, 1459–1470 (2015).

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation funded by the Korean Government (RS-2023-00228147).

AUTHOR CONTRIBUTIONS

J.W.C. conceived this study, co-wrote the manuscript, and analyzed the data. J.H.H. conceived this study, co-wrote the manuscript, and performed experiments. J.E.H. conceived this study, performed experiments, and analyzed the data. J.H.K. supervised the experiments. D.H.B. conceived this study, supervised the experiments, analyzed the data, and co-wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s12276-024-01164-8>.

Correspondence and requests for materials should be addressed to Ji Hyun Lee or Duhee Bang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024