Application Note

A miniaturized, semi-automated RNA-seq workflow for rapid lead selection and prioritization to streamline early drug development

Samuel Rulli[1], Stuart Tugendreich[2], Kamila Koprowska[3], Huw Rees[3], Krittika Mittal[4], Ke Xu[4], Niladri Basu[4]

[1] QIAGEN Sciences, Frederick, MD, USA

[2] QIAGEN Digital Insights, Redwood City, CA, USA

[3] SPT Labtech, Melbourn, Cambridgeshire, UK

[4] Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, Canada

Introduction

The primary goal of early screening in drug development is to identify candidates that deliver the intended biological effect while eliminating those with poor efficacy or unacceptable toxicity. Gene expression profiling provides early insights into how drug candidates influence cellular signaling and phenotype, helping to triage leads before downstream investment.

Standard RNA-seq approaches often prioritize polyadenylated transcripts, which can exclude most non-coding RNAs. This is typically achieved by enriching mRNA using oligo-dT capture methods or by priming the reverse transcription reaction with oligo-dT primers. Although this focuses on the protein-coding part of the genome, mRNA enrichment adds another step to the workflow, often requires higher starting input amounts and can miss non-polyadenylated non-coding RNAs[1].

For researchers aiming to uncover new biomarkers and molecular mechanisms, long non-coding RNAs are increasingly recognized as early markers of stress, apoptosis and other phenotypic changes. In addition, lncRNAs are known to encode biologically active ▷

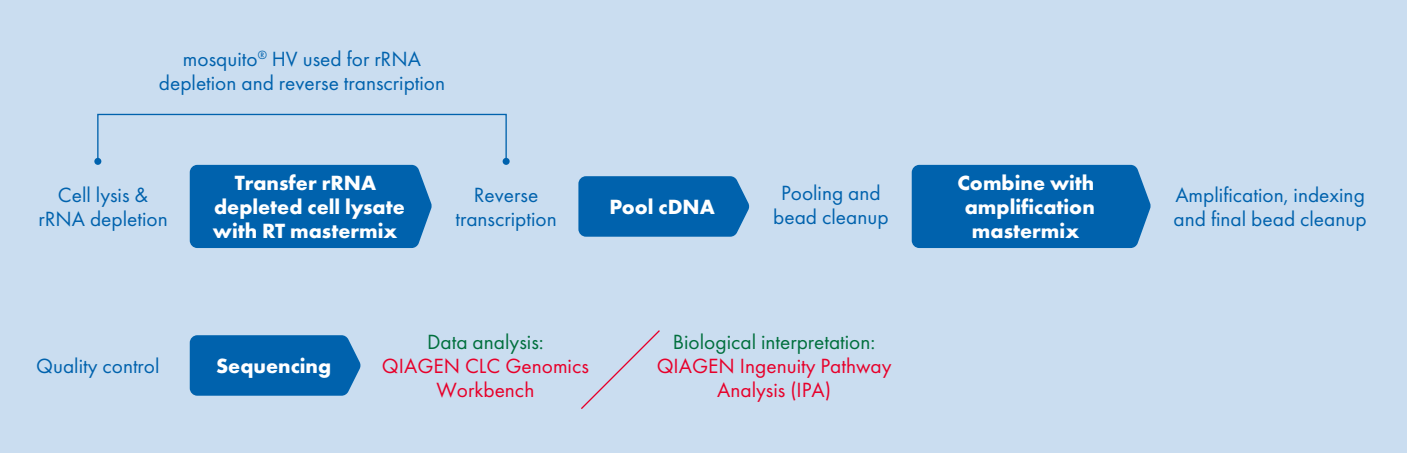

Figure 1. A high-throughput RNA-seq workflow that captures both coding and non-coding RNA. Following cell lysis, the RNA is fragmented and rRNA is suppressed. During cDNA synthesis, a unique sample ID is incorporated into each sample, enabling cDNA pooling. Pooled samples can then be indexed and amplified in a single step using PCR. Amplified libraries are purified, sequenced and then fed into ready-to-use pipelines for read mapping, gene counts, differential expression and Ingenuity® Pathway Analysis (IPA®) for biological insights.

micropeptides[2] that contribute to diverse cellular functions. As a result, the need to balance RNA-seq methods with different drug screening priorities often requires laboratories to maintain multiple RNA-seq library kits and workflows.

This Application Note describes a semi-automated, miniaturized platform for gene expression analysis that uses an off-the-shelf direct cell lysis RNA-seq assay and a nanoliter-scale liquid-handling device (Figure 1). This workflow can capture polyadenylated protein-coding RNA or both coding and non-coding RNA by simply changing the primer used during reverse transcription. The integrated QIAseq® FastSelect® ribosomal RNA blocking technology suppresses unwanted reads from both cytoplasmic and mitochondrial rRNAs. In addition, sample barcoding using UPXome technology during reverse transcription assigns a unique ID to each sample's cDNA, enabling cDNA pooling that simplifies and accelerates library construction. This platform allows research programs to generate high-quality data with greater flexibility, whether the focus is on protein-coding genes or on broader full transcriptome coverage to support informed decisions during drug development.

Materials and Methods

Sample preparation, including cell lysis

HEPG2 and Caco-2 cells were seeded at 35,000 cells per well in 96-well plates. After 24 h, cells were exposed to different concentrations of DMSO (solvent control) or H_2O_2 (to simulate oxidative stress) in serum-free DMEM. Following 24 h of exposure, cells were rinsed with 100 µL of cold PBS and lysed in 40 µL of 1x QIAseq UPX Cell Lysis reagent (13.3 µL lysis buffer, 23.3 µL nuclease-free water, 3.3 µL RNase inhibitor). Plates were incubated at room temperature with shaking for 10 min and then frozen until further processing.

Semi-automated RNA library preparation

RNA-seq libraries were prepared using the QIAseq UPXome RNA Library Kit with QIAseq FastSelect–rRNA HMR reagent to reduce cytoplasmic and mitochondrial ribosomal RNA. Reverse transcription can be primed with either oligo-dT to capture polyadenylated transcripts or random hexamers for full transcriptome coverage.

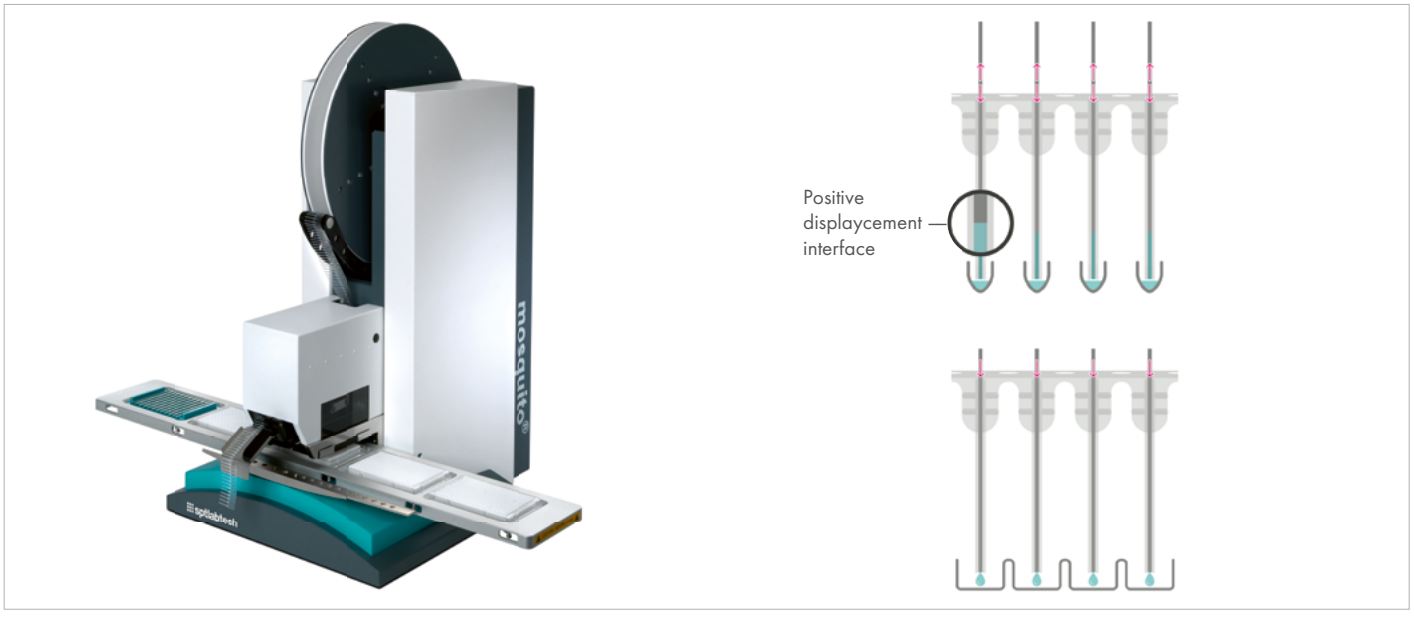

Figure 2. Miniaturization and automation of RNA-seq workflow on mosquito® HV genomics liquid handler. The system delivers precise nanoliter-to-microliter pipetting through positive displacement technology, reducing reagent use and sample input while ensuring consistent performance without liquid class optimization. High-density plate formats (96-, 384-, and 1536-well) support reproducible, large-scale library preparation with minimal waste. This setup lowers per-sample costs, improves scalability, and enables both small pilot studies and high-throughput screening within the same platform. Learn more here: www.sptlabtech.com/products/mosquito/mosquito-hv-genomics

Each pool contained 48 samples. Libraries were amplified and indexed with QIAseq UX Index Kits. The workflow was automated and miniaturized on the SPT Labtech mosquito HV genomics liquid handler to reduce reagent use and increase throughput (Figure 2 and Table 5).

Sequencing and data analysis

Libraries were sequenced using short-read RNA-seq. Reads were aligned to the human reference genome (GRCh38.noalt.106) and assigned to transcript features using QIAGEN® CLC Genomics Workbench. Mapping rates, exon coverage and protein-coding assignments were calculated for each condition. Differential expression analysis was performed using standard statistical filters (absolute fold change >2, adjusted p-value <0.05). Pathway enrichment and toxicity-related signatures were analyzed using QIAGEN IPA.

Results and Discussion

Performance of RNA-seq libraries using the automated workflow

For workflow verification, RNA-seq libraries were prepared using 0.5x volumes, both manually and via the semi-automated workflow, for Caco-2 cells. We chose to use a direct cell lysis method with assay miniaturization and to focus on polyadenylated RNAs by using an oligo-dT primer during reverse transcription. The RNA-seq mapping data demonstrated similar mapping rates between the manual and semi-automated workflows. By miniaturizing and automating specific steps, we achieved enhanced consistency in library preparation, reduced hands-on time and no loss in assay performance (Table 1).

RNA biotypes and read mapping for Caco-2 and HEPG2 cells using oligo-dT primer during treatments

The semi-automated, miniaturized workflow was used to assess changes in gene expression in Caco-2 and HEPG2 cells under different experimental conditions. The mapped RNAs (Table 2) were classified based on function. Caco-2 and HEPG2 cells showed a similar distribution of protein-coding, lncRNA and other RNA biotypes (Table 3).

The low mitochondrial and cytoplasmic rRNA levels are attributed to 3' sequencing using an anchored oligo-dT primer and to the suppression of internal priming of cytoplasmic and mitochondrial rRNA by FastSelect ribosomal RNA HMR technology.

Table 1. Percentage of mapped reads from RNA-seq libraries prepared from Caco-2 cells using mosquito HV genomics semi-automated workflow (0.5x volume) compared with manual (0.5x volume) preparation

	Percentage of mapped reads	
	Semi-automated using mosquito HV (0.5x volume)	Manual (0.5x volume)
Overall mapping	70–93	69–81
Mapped to genes	84–98	91–96
Mapped to introns	11–26	15–22
Mapped to exons	60–84	70–80
Mapped to intergenic region	2–13	4–9
Mapped to rRNA	0.04–0.89	0
Mapped to protein coding	68–77	74–77

Table 2. Summary of RNA-seq libraries covering coding and non-coding RNA obtained from Caco-2 and HEPG2 cell lines

	Caco-2	HEPG2
Number of samples	84	89
Average number of reads per sample	1,660,791	1,756,741
Percentage reads mapped in pairs	90	68
Percentage ignored reads (% wrong strand)	3	4
Percentage mapped to gene (intron + exon)	96	95
Percentage mapped to exon	78	75
Percentage mapped to intergenic region	4	5
Percentage protein coding	75	76

Table 3. Summary of RNA biotypes classified based on function

Biotype	Caco-2 cells (%)	HEPG2 cells (%)
Protein coding	75.3	76.3
lncRNA	1.9	1.6
Nonsense-mediated decay	7.3	7.0
Retained introns	8.5	7.8
snoRNA	1.5	1.7
Mitochondrial & cytoplasmic rRNA	0.1	0.2
Other RNA biotypes	5.4	5.4

Differential gene expression and IPA analysis

Treatment with hydrogen peroxide and DMSO generated measurable transcriptional changes between the treated samples and untreated cells. The differentially expressed genes and false discover rate (FDR) was calculated (Table 4) for each gene at different treatments. As expected, increasing doses of hydrogen peroxide and DMSO caused increased gene expression changes.

QIAGEN IPA was used to identify affected pathways and predict upstream regulators in response to increasing doses of DMSO or H_2O_2 for Caco-2 (Figure 5), and HEPG2 (not shown) as well as to compare cellular responses between the two studied treatments. IPA interprets observed gene expression changes by analyzing them in relation to a human-curated knowledge base of millions of literature findings. This enables the prediction of key drivers behind the gene expression alterations in treated cells and their downstream biological effects[4].

Table 4. Summary of differentially expressed genes with an FDR <0.05 for each concentration of DMSO and H_2O_2 across both cell lines used for IPA analysis

Caco-2		HEPG2		Caco-2		HEPG2	
DMSO (mM)	Differential genes	DMSO (mM)	Differential genes	H_2O_2 (mM)	Differential genes	H_2O_2 (mM)	Differential genes
0.08	2	0.04	225	0.025	5	0.01	276
0.15	33	0.09	102	0.05	33	0.03	56
0.3	5	0.17	50	0.10	20	0.06	110
0.6	63	0.34	95	0.42	16	0.12	50
1.2	1120	0.68	273	0.83	39	0.23	94
2.5	2612	1.37	725	1.67	1217	0.47	111
5	4311	2.73	2663	3.33	1275	0.94	175
10	1007	5.46	3076			1.88	1641

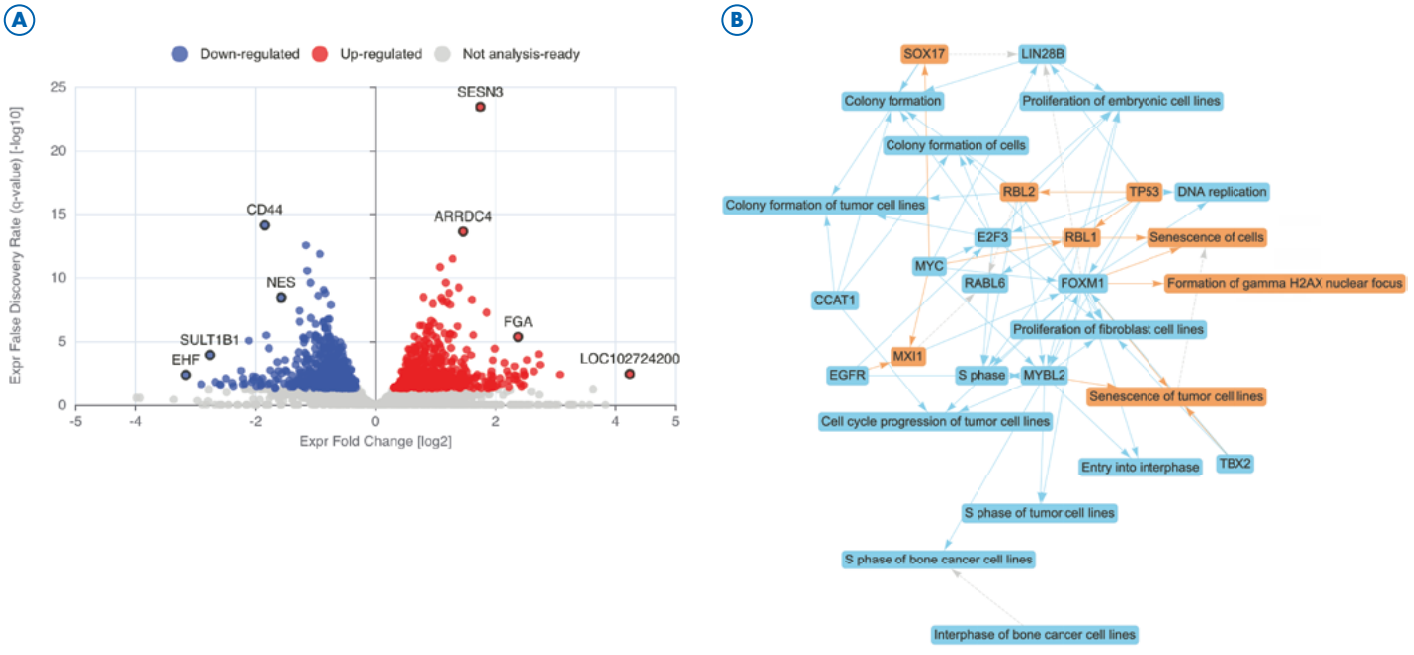

Figure 3. IPA results for Caco-2 cells treated with 1.2 mM DMSO. A The top portion of the IPA results summary page shows the dataset represented as a volcano plot. Red dots represent upregulated genes, while blue dots represent downregulated genes; both categories meet the analysis cutoffs. **B** The right portion shows a summary of the top predicted activities from the analysis as an interconnected network. Orange-colored nodes represent biological entities predicted to be activated, and blue nodes represent those predicted to be inhibited. For example, DMSO appears to inhibit cell cycle progression and increase senescence of Caco-2 cells. View the complete results here: https://analysis.ingenuity.com/ipw/fREdgXOB1hJ0WXTV2-eZ2A.

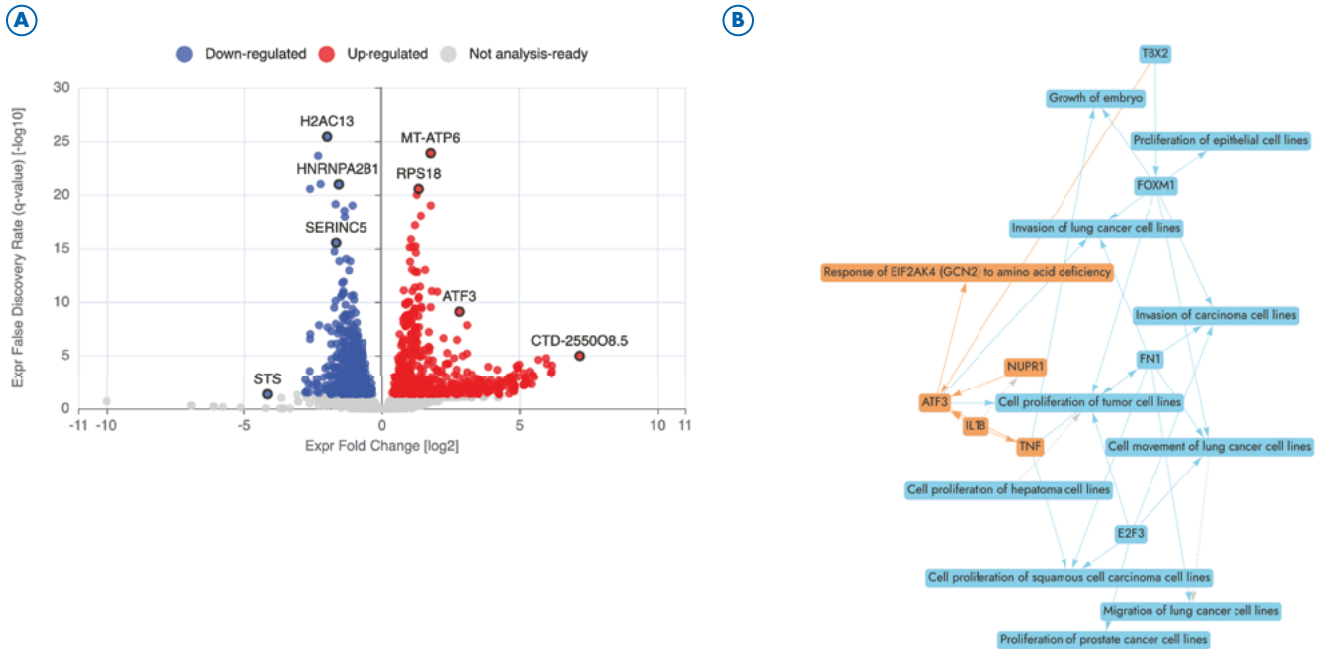

Figure 4. IPA results for Caco-2 cells treated with 3.3 mM H_2O_2. A The top portion of the IPA results summary page shows the dataset represented as a volcano plot. Red dots represent upregulated genes, while blue dots represent downregulated genes; both categories meet the analysis cutoffs. **B** The right portion shows a summary of the top predicted activities from the analysis as an interconnected network. Orange-colored nodes represent biological entities predicted to be activated, and blue nodes represent those predicted to be inhibited. H_2O_2 appears to increase several upstream regulators of inflammation, such as TNF and IL1B, and to decrease cell proliferation and cell movement. View the complete results here: https://analysis.ingenuity.com/ipw/_RNEsueZrPLlHzqWLAzPVQ.

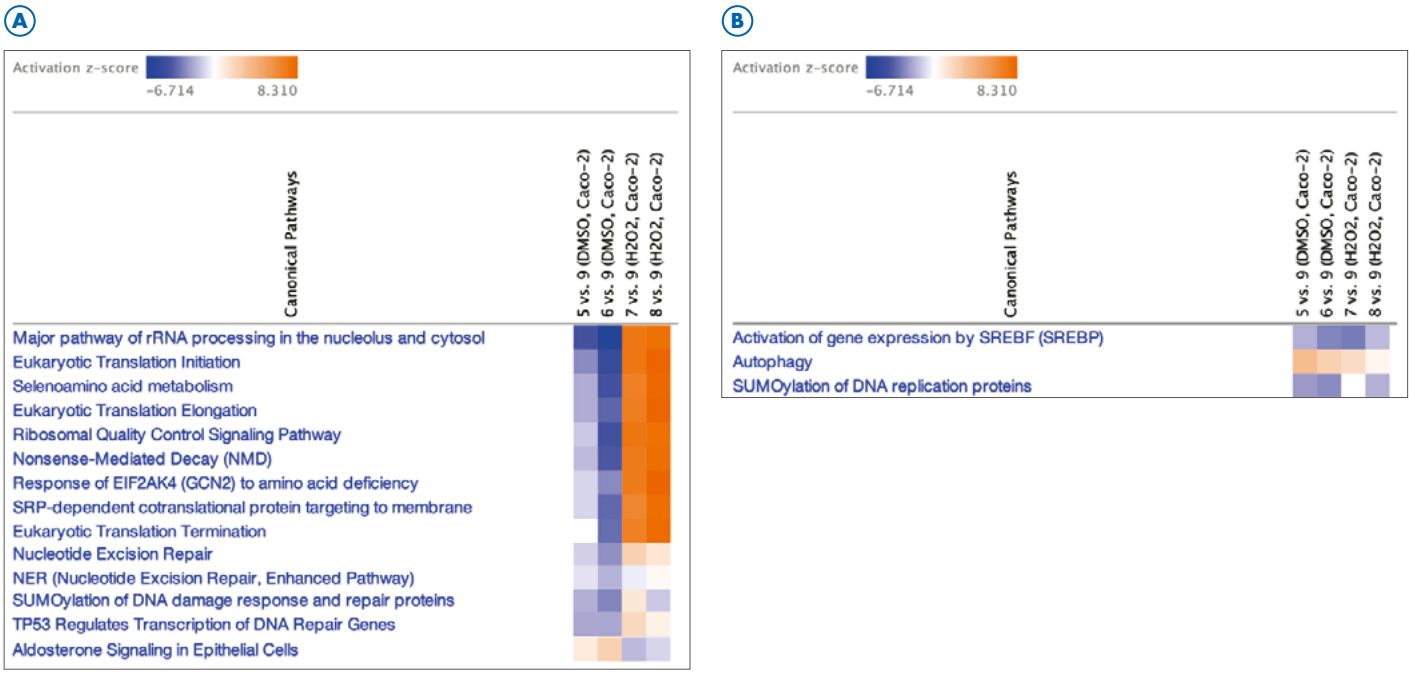

Figure 5. IPA identifies pathways predicted to be activated or inhibited in Caco-2 cells in response to DMSO and H_2O_2 treatments. A Pathways that are predicted to be inhibited (blue) or activated (orange) in response to the two chemical exposures. DMSO inhibits several pathways related to protein translation, whereas H2O2 activates them. **B** Examples of pathways that are similarly inhibited or activated by both compounds.

Sustainability, throughput and cost savings of standard, ultraplexed and miniaturized-ultraplex RNA-seq library construction

To assess the sustainability and efficiency of UPXome technology with miniaturization and automation, we compared standard RNA-seq library workflows with standard input to cDNA pooling (UPX technology) and to additional miniaturization. Several improvements to the standard workflow were accomplished to increase sample throughput while decreasing reagent and plastic usage. First, ultraplexing via UPXome technology, which introduces a sample ID into the cDNA reaction, enabled cDNA pooling. Pooling 24 samples into a single library reduced the amount of PCR reagents and plastics by 95%. Then, miniaturization of the reagents with the mosquito HV system reduced reagent volumes by half and allowed pooling of 48 samples per library. This lowered per-sample cost and plastic use compared to conventional manual workflows at standard volumes (Table 5). A 96-sample run consumed fewer plates, reservoirs and tips than manual library preparation, providing both economic and environmental benefits. The system supports processing in 96-, 384-, and 1536-well formats, making it suitable for both pilot studies and large-scale compound screens.

Table 5. Comparison of single-use plastic, reagents and time required for library preparation of 192 samples using a conventional RNA-seq kit, UPXome technology manually at 1x reaction volumes or with mosquito HV genomics at half reaction volumes

Step	Metric	Conventional RNA-seq	QIAseq UPXome technology	
		Manual	Manual (% change)	Semi-automated on mosquito HV (% change)
rRNA depletion and reverse transcription	Reaction volume	1x	1x	0.5x
	Reaction plates	2 x 96-well PCR plates	2 x 96-well PCR plates	Half 384-well PCR plate
	Reagent reservoir(s)	Variable, at least 2 reservoirs	2 reservoirs	Partial 384-well USD
	Tips	1188	288 (–76%)	248 (–79%) ▼
Pre-amplification pooling	Reaction plates	no	Yes, 24 samples per library pool ▼	Yes, 48 samples per library pool ▼
Amplification and bead clean-up	Reaction volume	1x	0.5x (–50%)	0.5x (–50%)
	Reaction plates	4 x 96-well PCR plates	0 (–100%)	0 (–100%)
	Tips	960	80 (–92%)	40(–96%)
	Assay reagent	24 µL	12 µL (–50%)	6 µL (–75%)
	Tubes	192	9 (–95%) ▼	4(–98%)
Quality check	Tips	192	9 (–95%) ▼	4(–98%)
	Tubes	96	8 (–92%) ▼	4(–96%)
Time (h)	Up to reverse transcription	24	16 (–33%) ▼▼	16 (–33%) ▼▼
	Amplification to the end	16	5 (–69%)	5 (–69%)

Conclusion

This miniaturized, semi-automated RNA-seq platform provides high-quality transcriptome data directly from cell lysates, maintaining coverage for protein coding RNA with minimal reads from cytoplasmic or mitochondrial ribosomal RNA. The workflow presented here can be further customized and scaled by making minor changes to oligos used during the reverse transcription reaction. By integrating automation, miniaturization and direct cell lysis, bulk RNA-seq can support robust differential expression and pathway analysis while reducing reagent use, plastic waste and per-sample cost. This approach would help increase the efficiency of drug candidate screening and prioritization, allowing more leads to be evaluated in parallel. The result is faster identification of promising candidates, reduced risk of advancing ineffective molecules and overall acceleration of early-stage drug development.

References

1. Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochim Biophys Acta.* 2016;1859(1):128–138.
2. Xiao Y, et al. Long non-coding RNA-encoded micropeptides: functions, mechanisms and implications. *Cell Death Discov.* 2024;10(1):450
3. Mittal K, Xu K, Rulli SJ Jr, Zhou G, Xia J, Basu N. TPD-seq: A high throughput RNA-seq method to derive transcriptomic points of departure from cell lines. *Toxicol In Vitro.* 2025;104:106001.
4. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics.* 2014;30(4):523–530.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

This Application Note is intended to share technical experience and workflow insights. Each party remains solely responsible for its own products and compliance with applicable laws and IP obligations.

Learn more about QIAseq UPXome RNA Library Kits: **www.qiagen.com/qiaseqlibkit**