

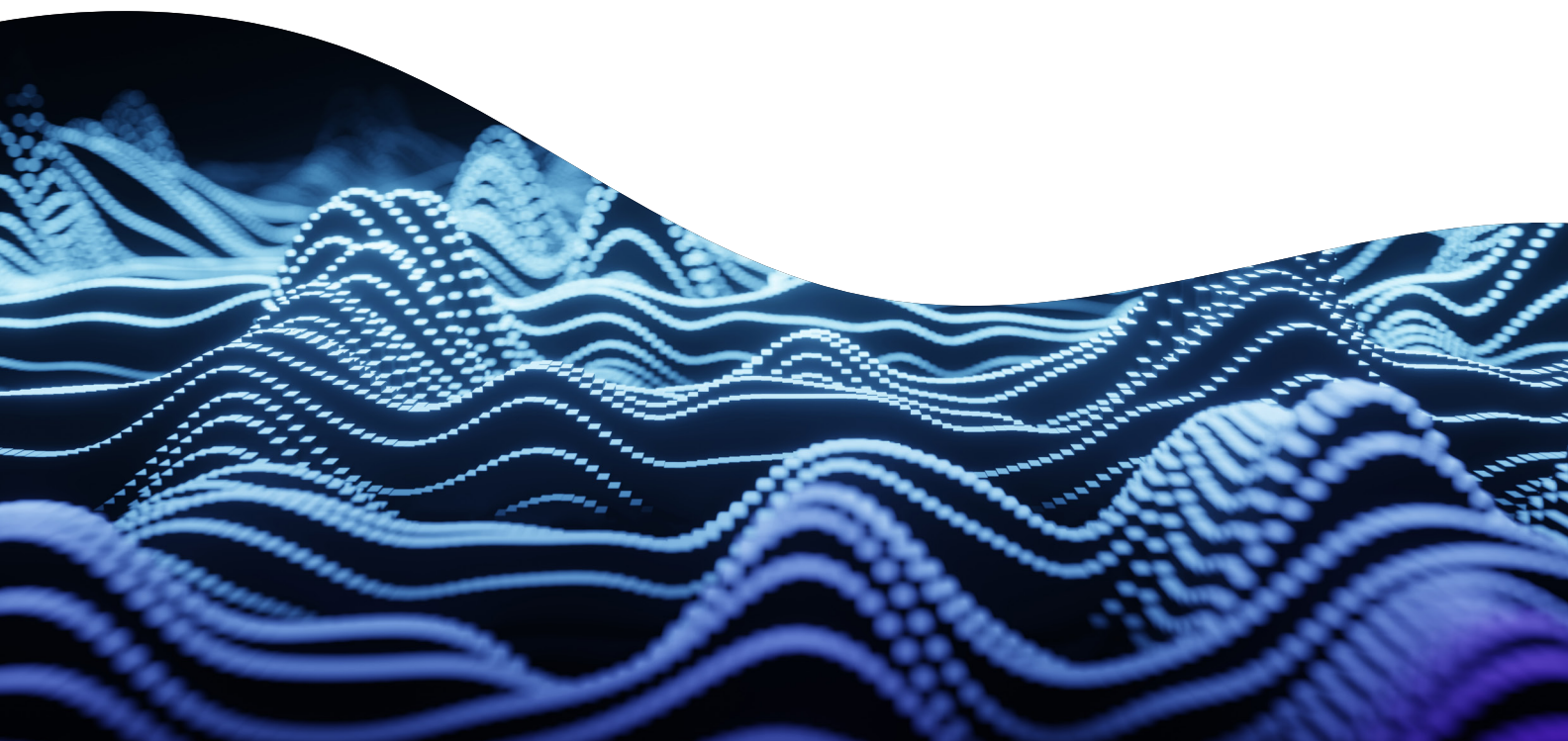
Benchmarking MERCURIUS™ DRUG-seq against conventional RNA-seq.

Introduction

RNA sequencing is widely used to characterize transcriptional responses to compounds, genetic perturbations, and other biological stimuli. Conventional mRNA-seq workflows generate reads across transcript bodies and involve RNA purification and separate library preparation for each sample. These requirements can become limiting as studies expand to larger numbers of compounds, experimental conditions, dose levels, or biological replicates.

MERCURIUS™ DRUG-seq is a high-throughput 3' mRNA sequencing workflow designed for scalable gene-expression profiling directly from cell lysates. Rather than sequencing across the full length of each transcript, DRUG-seq concentrates sequence information at the 3' end of polyadenylated transcripts to generate gene-level expression measurements. Extraction-free processing and early sample multiplexing allow samples to share downstream library-preparation steps. The 3' counting approach can also reduce sequencing requirements for gene-expression profiling¹.

Here, MERCURIUS™ DRUG-seq was benchmarked against the NEBNext® Ultra II mRNA sequencing workflow using Huh7 human hepatocellular carcinoma cells exposed to a TGF- β stimulus. Huh7 cells are a TGF- β -responsive model in which TGF- β /SMAD signaling and, under sustained exposure, epithelial-to-mesenchymal transition-associated transcriptional changes have been described^{2,3}.



Performance was evaluated across gene-body coverage, gene detection, gene-expression concordance, treatment-driven sample separation, differential expression, and KEGG pathway enrichment. The objective was to determine whether the screening-oriented DRUG-seq workflow could reproduce the principal gene-level and pathway-level responses identified using conventional RNA sequencing in this Huh7 perturbation model.

Methods

Experimental design

Huh7 hepatocellular carcinoma cells were treated with 10 ng/mL TGF- β for 48 h and analysed using either the MERCURIUS™ DRUG-seq or NEBNext® Ultra II mRNA sequencing workflow.

For MERCURIUS™ DRUG-seq, libraries were generated directly from cell lysates corresponding to approximately 10,000 cells per well. The experiment comprised 24 replicates per condition: 24 untreated and 24 TGF- β -treated samples. For NEBNext® Ultra II, libraries were prepared from 500 ng purified total RNA, with four replicates per condition. Libraries were sequenced on an Element Biosciences AVITI™ System using a paired-end read configuration, where Read 1 had 28 bp (barcode and UMI) and Read 2 had 100 bp (RNA sequence).

Sequencing depth normalization

MERCURIUS™ DRUG-seq generated approximately 7.8 million raw reads per sample. NEBNext® Ultra II mRNA generated approximately 16.3 million raw reads per sample. To compare the workflows on a common gene-assigned read-count basis, NEBNext® Ultra II data were down sampled to approximately 5.7 million ENSG-mapped reads per sample, matching the average ENSG-mapped depth obtained with MERCURIUS™ DRUG-seq. An ENSG-mapped read was defined as a sequencing read assignable to a specific Ensembl gene identifier.

Because MERCURIUS™ DRUG-seq is a 3' mRNA workflow, its reads are concentrated near the 3' end of polyadenylated transcripts rather than distributed across the full transcript body. Matching by ENSG-mapped reads therefore provided a pragmatic common basis for gene-level comparison but does not imply equivalent transcript-body coverage or identical information content per read (Figure 1).

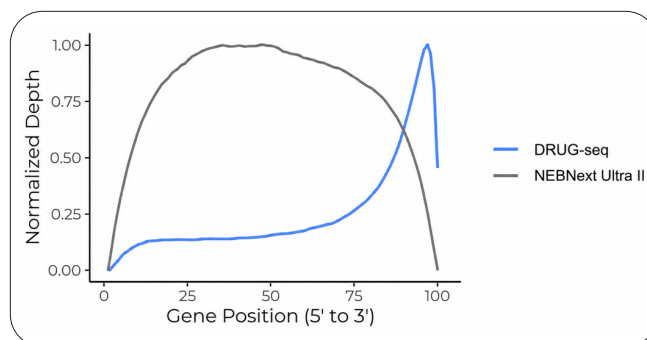


Figure 1: Gene-body coverage profiles reflect the expected difference between the workflows. MERCURIUS™ DRUG-seq showed strong enrichment toward the 3' end of genes, whereas NEBNext® Ultra II generated broader coverage across the gene body.

Gene detection and expression stratification

Gene detection was evaluated in two complementary analyses. First, the number of detected genes was compared across raw-read depths from 0.1 to 7 million reads per sample. Second, at matched ENSG-mapped depth, genes were stratified by counts per million reads (CPM) as low (>0 to <1 CPM), intermediate (≥ 1 to <4 CPM), or high (≥ 4 CPM). A gene was considered detected when CPM >0 .

Correlation and differential expression

Gene-expression measurements were compared separately in untreated and TGF- β -treated samples using genes detected by both workflows at >1 CPM. Pearson correlation was used to compare gene-expression measurements between workflows.

Differentially expressed genes were defined using an adjusted P value <0.01 and fold-change magnitude >1.5 (equivalent to $|\log_2 \text{fold change}| >0.585$). Overlap between workflows and correlation of fold-change estimates were subsequently assessed.

Higher-level biological analysis

Principal component analysis (PCA) was used to assess sample reproducibility and treatment-driven separation. KEGG pathway enrichment analysis⁴ was performed on the differential-expression results to compare the biological pathways recovered by the two workflows.

Results

Gene detection increases with sequencing depth in both workflows

Gene detection increased with sequencing depth in both workflows. NEBNext® Ultra II had a higher median number of detected genes at every raw-read depth tested. At the common gene-assigned read count used for downstream analyses, both workflows captured broad gene-expression profiles suitable for comparison of treatment effects (Figure 2).

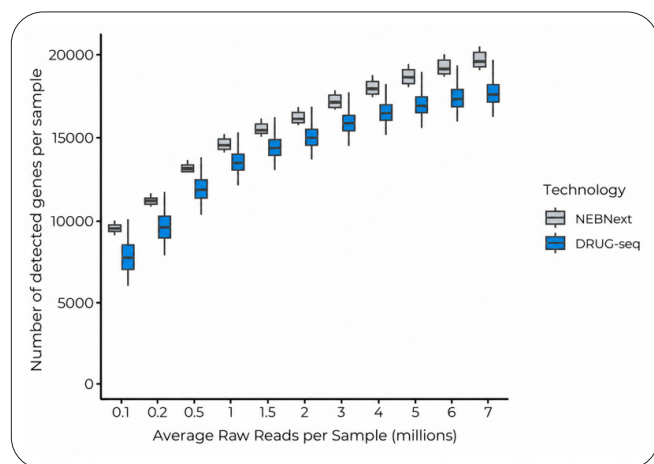


Figure 2: The number of genes detected per sample was compared between MERCURIUS™ DRUG-seq and NEBNext® Ultra II across a range of raw sequencing depths. Gene detection increased with read depth in both workflows. NEBNext® Ultra II had a higher median number of detected genes at each tested depth.

NEBNext® Ultra II contained more detected genes in the low- and high-CPM bins, whereas MERCURIUS™ DRUG-seq contained more genes in the intermediate-CPM bin (Figure 3). Because these categories are defined from each workflow's own CPM values, this comparison describes the distribution of detected genes across expression bins and should not by itself be interpreted as a direct measurement of analytical sensitivity.

Despite these differences, the overall numbers of detected genes remained of similar order between workflows.

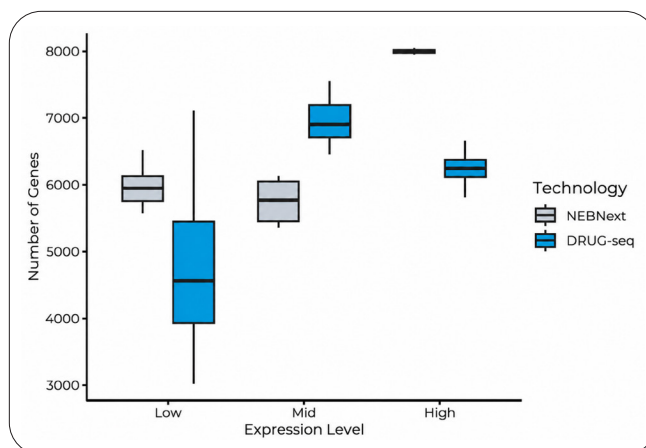


Figure 3: Gene detection was compared at matched gene-assigned read depth after stratifying genes by expression level: low (>0 to <1 CPM), intermediate (≥ 1 to <4 CPM), and high (≥ 4 CPM). Because the bins were defined from CPM values within each workflow, this analysis is descriptive and should not be interpreted by itself as a direct analytical-sensitivity measurement.

Gene-expression measurements are strongly correlated between workflows

For genes expressed above 1 CPM and detected by both workflows, gene-level expression was correlated between workflows (Figure 4). These results indicate that, for commonly detected genes, MERCURIUS™ DRUG-seq showed relative expression patterns broadly comparable to those measured with NEBNext® Ultra II.

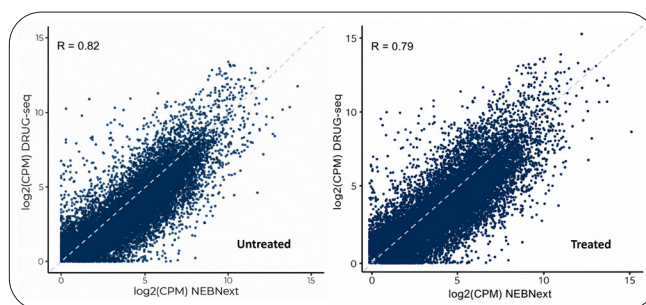


Figure 4: Gene-expression measurements were compared for genes detected by both workflows at >1 CPM in untreated and TGF- β -treated conditions. Expression values were strongly correlated between MERCURIUS™ DRUG-seq and NEBNext® Ultra II in both untreated samples ($R = 0.82$) and 10 ng/mL TGF- β -treated samples ($R = 0.79$).

Differential expression overlaps between workflows

Using an adjusted P value threshold of <0.01 and a fold-change magnitude threshold of >1.5 , the two workflows identified substantially overlapping sets of differentially expressed genes. A total of 4,740 differentially expressed genes were common to both workflows, representing 52.7 % of the union of genes called differentially expressed by either workflow (Figure 5, left).

The workflows were analysed with different numbers of replicates per condition (24 for MERCURIUS™ DRUG-seq and four for NEBNext® Ultra II). This imbalance can affect differential-expression power and therefore complicates direct comparison of the numbers of workflow-specific differentially expressed genes.

For these shared differentially expressed genes, $\log_2$ fold-change estimates were strongly correlated between workflows ($R = 0.96$; Figure 5, right). This comparison is restricted to genes called differentially expressed by both workflows.

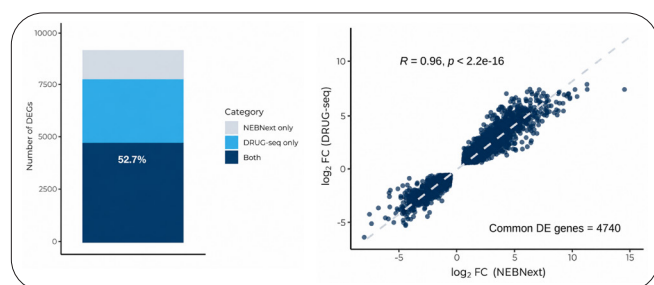


Figure 5: Left: Differential gene expression between untreated and 10 ng/mL TGF- β -treated Huh7 cells was evaluated using an adjusted P value <0.01 and fold-change magnitude >1.5 . A substantial proportion of differentially expressed genes was identified by both workflows, with 4,740 genes shared between the two datasets. The larger MERCURIUS™ DRUG-seq-only set should be interpreted in the context of the unequal sample numbers and should not be attributed to workflow sensitivity alone. Right: For differentially expressed genes identified by both workflows, $\log_2$ fold-change estimates were strongly correlated within this shared gene set ($R = 0.96$).

Both workflows separate untreated and TGF- β -treated samples

PCA showed clustering of replicates within each workflow and clear separation between untreated and TGF- β -treated samples. PC1 and PC2 explained 45.99 % and 33.98 % of the variance, respectively. Samples separated by workflow mainly along PC1 and by treatment mainly along PC2 (Figure 6).

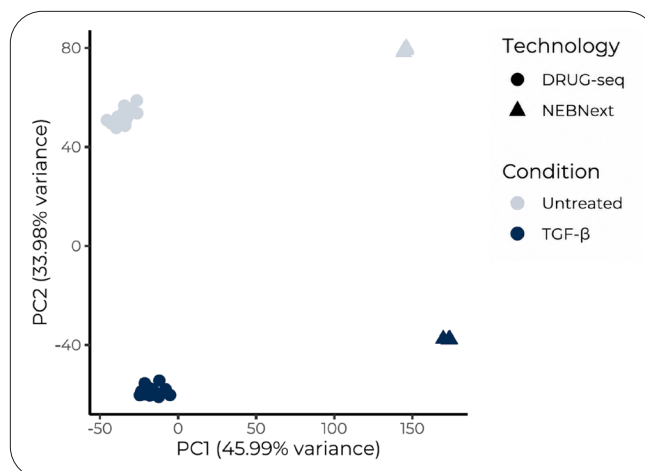


Figure 6: Principal component analysis was used to visualize sample clustering and treatment-associated transcriptional differences. Replicates clustered consistently within each experimental condition, while untreated and 10 ng/mL TGF- β -treated samples separated clearly with both workflows. PC1 and PC2 accounted for 45.99 % and 33.98 % of the variance, respectively.

Both workflows identify enrichment of the TGF- β signaling pathway

KEGG pathway enrichment analysis identified pathways enriched in both datasets, including the TGF- β signaling pathway (Figure 7). Other pathways were classified as enriched by only one workflow. The shared TGF- β pathway result is consistent with the applied perturbation.

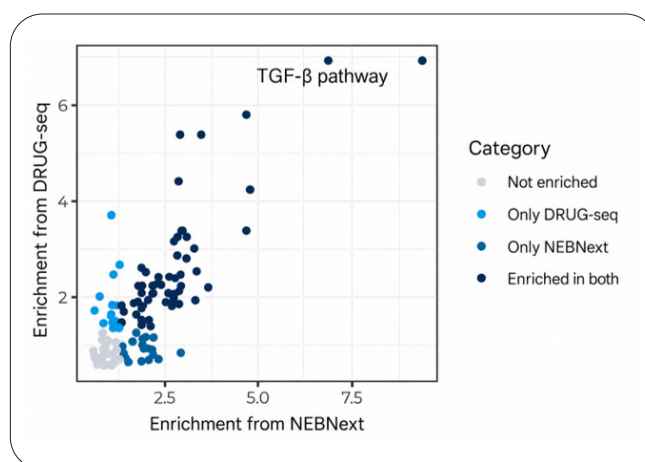


Figure 7: KEGG pathway enrichment was compared using the differential-expression results from MERCURIUS™ DRUG-seq and NEBNext® Ultra II. Each point represents a pathway; colours indicate whether it was classified as enriched by either workflow, both workflows, or neither. The TGF- β signalling pathway was identified as enriched in both datasets.

Conclusion

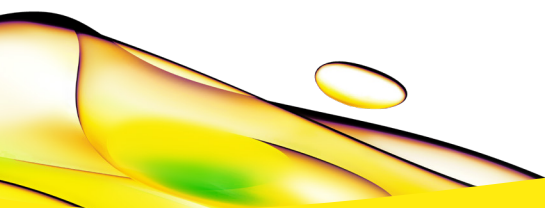
In Huh7 cells treated with 10 ng/mL TGF- β for 48 h, MERCURIUS™ DRUG-seq captured treatment-associated transcriptional changes using an extraction-free 3' mRNA-seq workflow. NEBNext® Ultra II showed broader gene-body coverage and a higher median number of detected genes at each raw-read depth tested. Gene-expression measurements were correlated for genes detected by both workflows.

Shared differential-expression results and TGF- β pathway enrichment support the use of MERCURIUS™ DRUG-seq for gene-level profiling in this model. Unequal replicate numbers limit comparisons of workflow-specific differential-expression calls.

Together, these results illustrate the balance between gene-level detection, transcript-body coverage, and scalability offered by MERCURIUS™ DRUG-seq. For applications in which throughput, replication, differential-expression signatures, and pathway-level responses are the primary objectives, MERCURIUS™ DRUG-seq provides an efficient and highly scalable 3' mRNA-seq strategy.

References

1. Ye, C., et al. (2018). DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat Commun.* 9:4307. doi:10.1038/s41467-018-06500-x.
2. Liang J., et al. (2020). Comprehensive analysis of TGF- β -induced mRNAs and ncRNAs in hepatocellular carcinoma. *Aging (Albany NY)*. 12(19):19399-19420. doi:10.18632/aging.103826.
3. Srivastava, A., et al. (2023) Transforming growth factor- β mediated regulation of epigenome is required for epithelial to mesenchymal transition associated features in liver cancer cells. *Heliyon*. 9(4):e14665. doi:10.1016/j.heliyon.2023.e14665.
4. Kanehisa, M., et al. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49(D1):D545-D551. doi: 10.1093/nar/gkaa970.



revvity