

Benchmarking the Performance of MERCURIUS™ Total DRUG-seq Sequenced on AVITI™ Systems with Paired-End Alignment

Abstract

The paper demonstrates how paired-end sequencing of multiplexed RNA-seq libraries that is currently only possible on AVITI™ systems, boosts transcript detection rate while uncovering physiologically-relevant promoter switching events.

Highlights

- Robust detection of 15,000 protein-coding genes at only 1.5 million reads per sample
- 40% higher transcript detection rate when aligning paired-end data versus single-end data
- 45% more differentially expressed transcripts detected with paired-end alignment after TGF- β treatment
- Sensitive detection of promoter switching events in paired-end data following TGF- β treatment

This paper was done in collaboration with:



Introduction

High-throughput screening drives compound discovery and development in the pharmaceutical, agritech, and cosmetics sectors. It is also becoming essential for generating the large-scale datasets needed to train and refine the next generation of artificial intelligence models.

Until recently, however, transcriptome-wide high-throughput screening was largely out of reach. Most gene expression profiling methods lack the scalability required to deliver cost-effective whole-transcriptome insights across hundreds or thousands of experimental conditions. As a result, transcriptomic screens using RNA sequencing (RNA-seq) are typically limited to relatively few samples, while more scalable but biased alternatives, like targeted gene panels or 3' mRNA-seq, miss important features like transcript isoforms, alternative splicing events, and novel transcripts.¹⁻³

To overcome these limitations, Alithea Genomics developed MERCURIUS™ Total DRUG-seq, a novel, scalable and massively multiplexed bulk RNA-seq technology. The approach uses sample and transcript barcoding to provide full-length transcriptome-wide data for both protein-coding and non-coding transcripts across up to 384 samples simultaneously.

On Illumina sequencers, MERCURIUS™ Total DRUG-seq libraries are typically sequenced in paired-end read configuration, in which read 1 is

used for the barcode and the UMI, while read 2 is used to read the transcript sequence. True paired-end sequencing is therefore not possible because Illumina systems do not allow for sequencing through the poly-A tail. Unlike Illumina's chemistry, Element Biosciences' AVITI™ sequencing chemistry can conveniently sequence through the poly-A tail and into the transcript sequence to provide true paired end sequencing of multiplexed RNA-seq libraries for the first time.

To investigate the advantages of this approach, Alithea Genomics aimed to benchmark paired-end (PE) versus single-end (SE) read alignments to evaluate whether PE alignment enhances the detection of expressed genes, transcript isoforms, and differentially expressed genes in a representative study of 96 samples treated with different concentrations of TGF-β. They also assessed the sensitivity to detect promoter switching events.

Overall, MERCURIUS™ Total DRUG-seq, combined with the AVITI™ sequencing platform, now enables scalable, full-length, transcriptome-wide sequencing optimized for high-throughput screening, providing crucial insights into drug response at a scale not possible with other technologies.



Methods

Compound Treatment

Huh7 hepatocarcinoma cells were seeded into a 96-well plate at a density of 10,000 cells per well. Twenty hours after seeding, cells were treated with 0.1 ng/mL, 1 ng/mL, 10 ng/mL TGF- β , or left untreated, for a total of 24 samples per dosage. After 48 hours of treatment, cells were washed with PBS, frozen, and then lysed using the proprietary MERCURIUS™ cell lysis buffer before library preparation.

Library Preparation

Libraries were prepared with the MERCURIUS™ Total DRUG-seq protocol by Alithea Genomics. The technology is RNA-extraction-free, with libraries generated directly from cell lysates. Extensive sample multiplexing is possible due to the optimized per-well sample barcodes, which enable the multiplexing of 96 samples into one tube at the reverse transcription stage, early in the pipeline. Unique molecular identifiers (UMIs) are also included to provide robust gene expression readouts by mitigating the effect of PCR duplicates introduced during amplification. The preparation included unique dual indices necessary for sequencing on the AVITI™ platform. The standard AVITI™ preparation protocol was followed. The standard AVITI™ preparation protocol was followed.



Sequencing

Following library preparation and quality control checks, the library was amplified with the Adept™ Rapid PCR-Plus Kit (Cat. #830-00018) and sequenced using Element's AVITI™ 2x150 Sequencing Kit Cloudbreak High Output (Cat. #860-00003) using Expert Mode HD and a custom recipe to support library diversity considerations. The library was sequenced with an 8bp Index 1 and Index 2 and asymmetric PE insert read lengths of 201bp and 101bp for Read 1 and Read 2. This resulted in a total of 968,120,838 reads for the run, with an average of 10,029,810 reads per sample after demultiplexing. 'Single-end' data was generated from this original PE data by computationally removing one read from each pair.



Results

Robust Gene Detection with Paired-End and Single-End Reads

To evaluate the benefit of aligning PE reads versus SE reads after sequencing of the MERCURIUS™ Total DRUG-seq library with Element's AVITIT™ platform, we first assessed the total number of protein-coding genes detected after downsampling to between 0.1 and seven million reads per sample (Fig. 1). At all downsampled read depths, a similar number of genes were detected in SE and PE data. Even at ultra-low read depths of around 1.5 million reads per sample, MERCURIUS™ Total DRUG-seq robustly detected around 15,000 genes in SE and PE data.

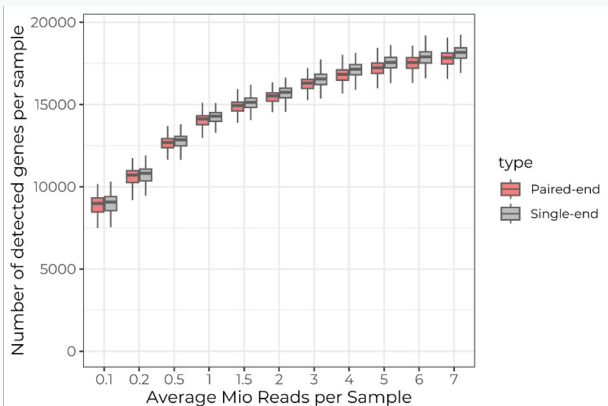


Figure 1. Boxplot indicating the number of genes detected per sample with PE and SE data at downsampled read depths.

Higher Transcript Detection with Paired-End Reads

As MERCURIUS™ Total DRUG-seq is a full-length protocol that provides reads along the whole transcript length and is not restricted to protein-coding genes, the total amount of detected transcripts was assessed. At the same sequencing depth, a consistently higher detection rate of transcripts was observed in PE data compared to SE data (Fig. 2). For instance, at a depth of only three million reads per sample, SE data detected approximately 45,000 transcripts, whereas around 63,000 were detected with PE data; a 40% increase.

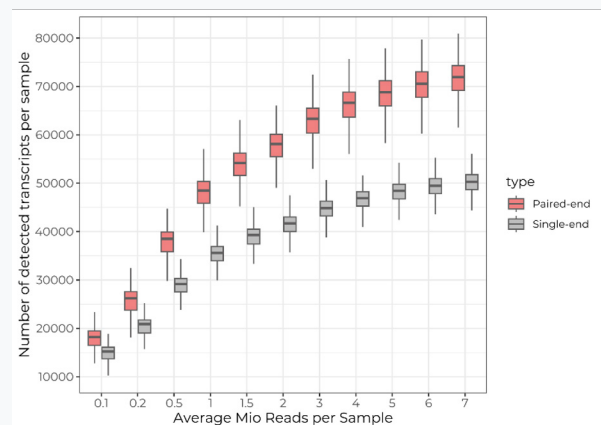


Figure 2. Boxplot indicating the number of detected transcripts per sample at different downsampled read depths.

Paired-end sequencing boosts the detection of differentially expressed genes

We next assessed the ability of MERCURIUS™ Total DRUG-seq to detect differentially expressed transcripts in PE and SE data from the 24 samples treated with 1 ng/mL TGF- β versus untreated samples. At the same sequencing depth of three million reads per sample, we observed an approximate 45% increase in the number of differentially expressed transcripts when using PE data (Fig. 3). Around 1900 upregulated genes and 2300 downregulated genes were detected in the PE data, versus 1200 upregulated genes and 1700 downregulated genes in the SE data.

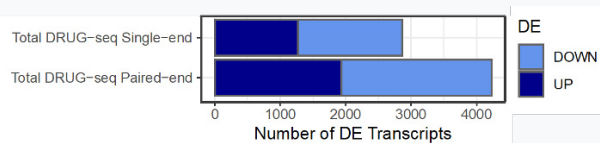


Figure 3. Bar chart indicating the total number of significantly differentially expressed transcripts following treatment with 1 ng/mL TGF- β for SE and PE data.

Paired-end MERCURIUS™ Total DRUG-seq detects promoter switching events

As MERCURIUS™ Total DRUG-seq is a full-length technology, it enables the fine-grained detection of isoform expression and facilitates the identification of promoter switching events. Promoter switching refers to

the dynamic shift in transcriptional start sites within a gene and is a crucial mechanism in drug discovery that potentially impacts therapeutic outcomes by influencing RNA stability, translation efficiency, and ultimately the protein product.⁴ As some diseases like cancer exhibit aberrant promoter switching, identifying these changes can lead to the development of drugs that can modulate these switches for targeted gene regulation and the development of novel therapies.⁴ We therefore used the proActiv algorithm to detect significant occurrences of promoter switching following compound treatment. In samples treated with 1 ng/mL and 10 ng/mL TGF- β , 19 and 22 significant switching events were detected, respectively (Fig. 4).

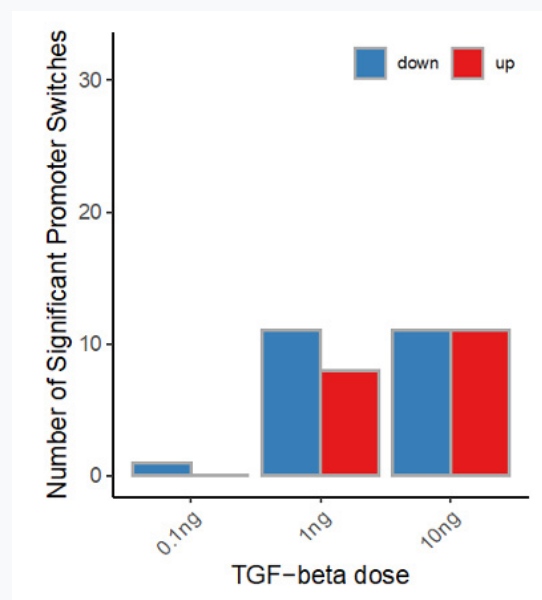
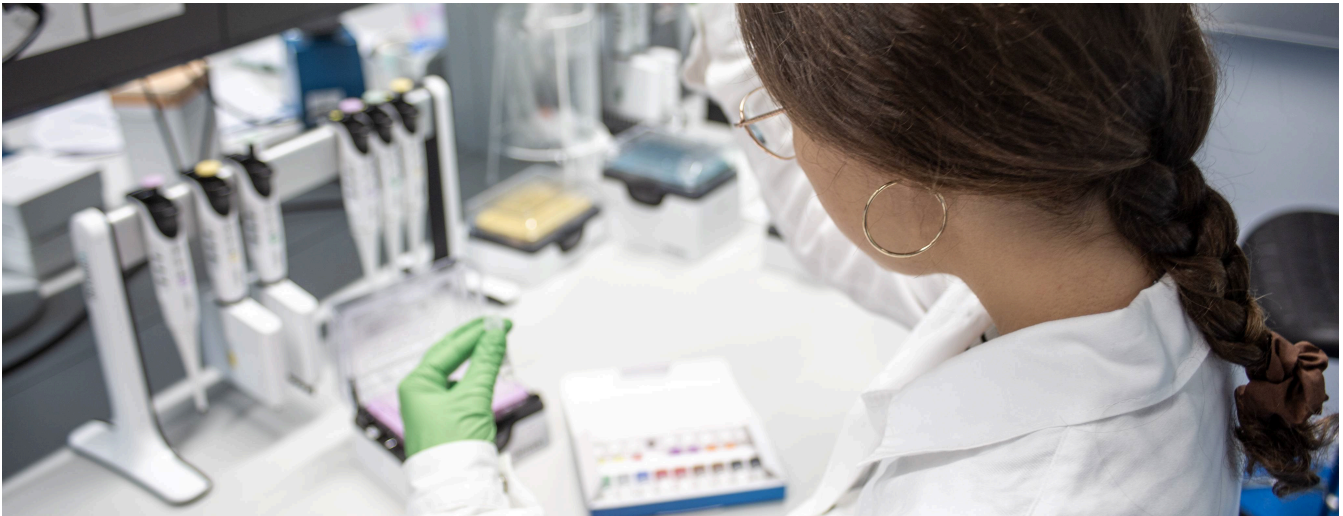


Figure 4. Bar chart indicating the number of significant promoter switching events detected in paired-end data following treatment with TGF- β .



The most significant promoter switch was detected for the HNF4a gene, a transcription factor whose transcript isoforms and their interplay with cofactors contribute to the complexity of gene regulation by HNF4a and its role in cancer.⁵ In paired-end data, the short isoform was active in untreated Huh7 cells (driven by 'promoter 2') but was inhibited and switched to the long isoform (driven by 'promoter 1') upon TGF- β treatment (Fig. 5).

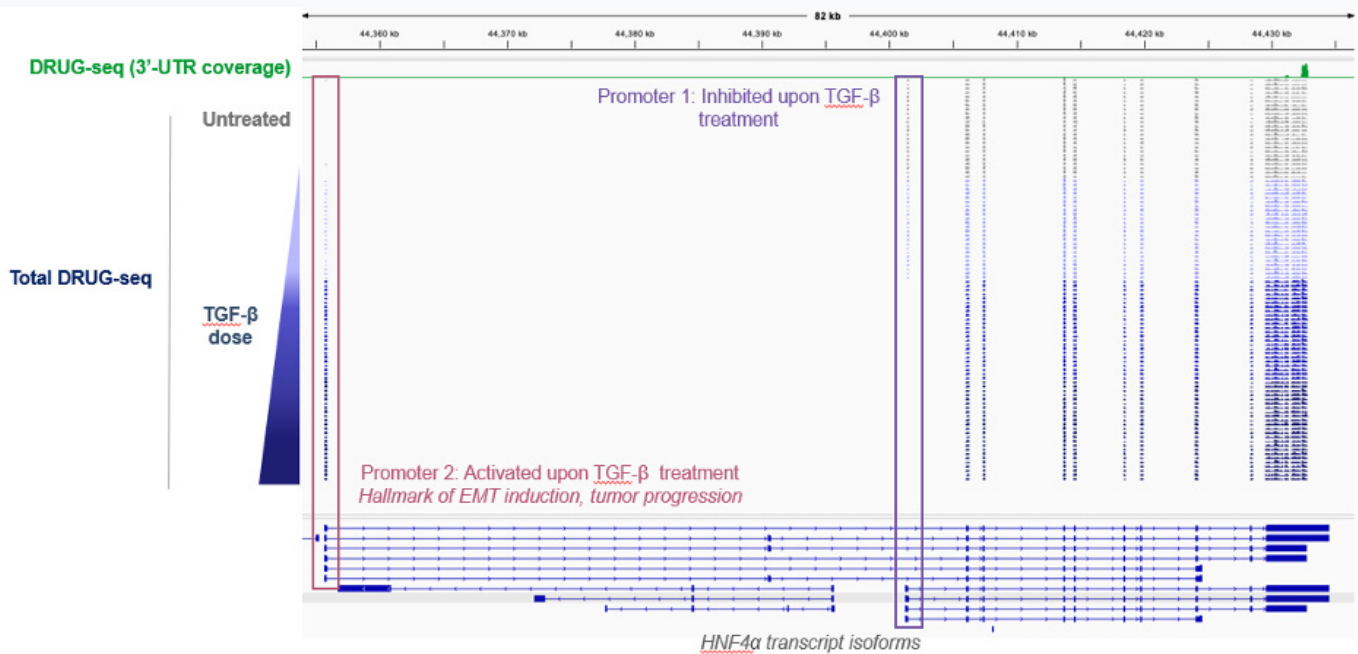


Figure 5. Integrative Genomics Viewer showing MERCURIUS™ Total DRUG-seq paired-end reads aligned to different full-length HNF4a transcript isoforms. Boxes indicate the locations of promoters 1 and 2, which switch following TGF- β treatment. Nine representative tracks are shown for each condition, with one track indicating the read distribution from 3' MERCURIUS™ DRUG-seq technology.

Results

MERCURIUS™ Total DRUG-seq is a highly scalable, sensitive, and full-length technology that offers optimal performance for quantifying whole transcript isoforms following large-scale compound treatment, utilizing paired-end sequencing with Element's AVITI™ technology. The extensive sample multiplexing made possible by MERCURIUS™ Total DRUG-seq, combined with the robust and flexible sequencing capabilities of the AVITI™ platform, now enables high-throughput screening studies to assess the transcriptomic effects of perturbations on over 60,000 transcripts at a sequencing depth of only three million reads per sample in a highly cost-effective manner.

The broad transcript coverage provided by paired-end alignment enables users to detect significantly more physiologically relevant transcript isoforms and promoter switches, which may help uncover crucial molecular mechanisms, novel biomarkers, or off-target effects of compounds that are not possible with other methods.

Overall, the combined use of MERCURIUS™ Total DRUG-seq and the AVITI™ platform is highly beneficial for accelerating large-scale studies in drug discovery, toxicology, agritech, and beyond.

References

1. Alpern, Daniel, et al. "BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing." *Genome biology* 20 (2019): 1-15.
2. Subramanian, Aravind, et al. "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles." *Cell* 171.6 (2017): 1437-1452.
3. Yeakley, Joanne M., et al. "A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling." *PloS one* 12.5 (2017): e0178302.
4. Wragg, Joseph W., et al. "Intra-promoter switch of transcription initiation sites in proliferation signaling-dependent RNA metabolism." *Nature Structural & Molecular Biology* 30.12 (2023): 1970-1984.
5. Qu, Ningxin, et al. "Hepatocyte nuclear factor 4 a (HNF4a): A perspective in cancer." *Biomedicine & Pharmacotherapy* 169 (2023): 115923.

