MERCURIUS™ 1536 DRUG-seq: Screening Scale Transcriptomics as a Foundation for AI-Enabled Drug Discovery

10,000+
genes detected at 200k reads/sample

400
cells /well

2.5 hr
hands-on workflow

Key Takeaways

MERCURIUS™ 1536 DRUG-seq provides reproducible, screening-scale transcriptomics, enabling biopharma to build structured, proprietary foundational data moats for AI-enabled drug discovery.

- Screening-scale, automation-compatible bulk 3' mRNA-seq: RNA-extraction-free, 2.5-hour hands-on workflow integrates directly into existing HTS pipelines to complement existing assays performed in 1536WP
- Sensitive gene detection at ultra-low read depths: Over 10,000 genes at 200K reads per sample detected consistently across five cell types
- Miniaturized reactions: Low recommended cell density of 400 cells per well reduces culture costs without compromising gene detection
- Plate-wide uniformity: Uniform read depth across all 1536 wells with minimal edge effects to maximize screening capacity
- AI-scale cost efficiency: per-sample costs aligned with other screening technologies, enabling prospective transcriptomic data moat generation at true screening scale

Introduction

A major biopharma-wide bottleneck in AI-enabled drug discovery is the lack of standardized, biologically native datasets of sufficient size, structure, and depth to train in silico models that accurately predict compound mechanisms of action, identify early activation of toxicity-associated pathways, and help to progress compounds with greater confidence (1).

Drug discovery teams can now employ RNA-extraction-free bulk transcriptomic methods such as DRUG-seq to provide transcriptome-wide 3' mRNA readouts in 96- and 384-well plate formats (2,3). While this format is suited to low- to mid-scale screens, this technology hasn't yet been implemented in the 1536-well plates required by high-throughput screening pipelines to generate true AI-scale foundational datasets that capture how diverse compounds affect cellular biology across doses, models, and conditions.

Transcriptomic data reveal pathway activation, gene regulatory changes, and off-target effects that morphological profiling methods cannot. Implementing transcriptomics in 1536-well plates stands to complement existing AI-scale morphological approaches, such as Cell Painting, by providing the molecular resolution underlying phenotypic changes (4,5). This multi-level view was previously challenging to achieve at screening scale, thus limiting AI models to a partial view of perturbational effects and constraining the robustness of in silico training, mechanism-of-action clustering, bioactivity modeling, and toxicity detection (4,6,7).
Here, we present the technical performance of MERCURIUS™ 1536 DRUG-seq across five cell types in three independent experiments.

The data indicate that MERCURIUS™ 1536 DRUG-seq provides biopharma with reliable, automation-compatible detection of >10,000 genes from 1536 cell lysates simultaneously, at depths as low as approximately 200K reads per sample. We demonstrate that the technology is reproducible across independent experiments, while reducing the per-sample cost of generating transcriptomes to a level necessary for AI-scale foundational datasets.

This now enables biopharma and AI-native CROs to generate prospective, standardized transcriptome-wide perturbation data at true screening scale, so that every screen builds a compounding foundational data moat, helps overcome the AI training data bottleneck, and allows in silico programs to continue to reach their promised potential.

Methods

To assess the performance of MERCURIUS™ 1536 DRUG-seq across different settings and cell types, we conducted three independent experiments. The overall experimental workflow was broadly similar in both instances as per the user guide (Fig. 1; Table 1).

Briefly, the lysis and reverse transcription master mix is dispensed into four barcoded oligo-dT 384-well plates, for a total of 1536 unique barcodes, using a standard liquid handling robot, such as firefly (SPT Labtech) or Bravo (Agilent). Oil is then added to the 1536-well plate using a nanodispenser or liquid handling robot, followed by the transfer of master mix and oligos from 384-well plates to the quadrants of the 1536-well plate. Oil serves to prevent evaporation and cross-contamination. The plate is then centrifuged and incubated at 37 °C in a cell incubator or oven for the reverse transcription reaction.

As MERCURIUS™ 1536 DRUG-seq is a 3' bulk mRNA-seq method, sample barcodes and unique molecular identifiers are added to the 3' poly(A)-tail of mRNA molecules during reverse transcription. The barcoded samples are then pooled in a single reservoir by upside-down centrifugation or pipetting, so that subsequent steps of the reaction occur in a single tube to reduce inter-sample variability and consumable use. The cDNA from this tube is column-purified, and the standard MERCURIUS™ DRUG-seq protocol then converts the single-strand cDNA into a barcoded library, ready for sequencing.

The entire protocol for a full 1536-sample library preparation requires approximately 2.5 hours of hands-on time, depending on the automation setup.

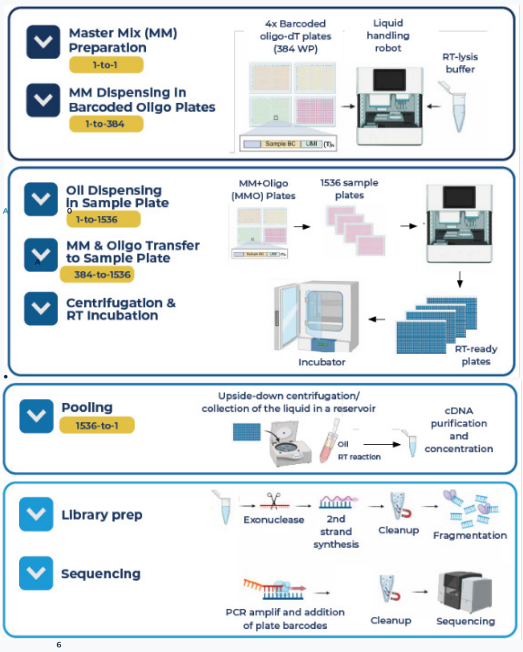


Figure 1. Schematic overview of the MERCURIUS™ 1536 DRUG-seq workflow. For more in-depth protocol information, please see our user guide.

	Experiment 1	Experiment 2	Experiment 3
Cell type	HEPG2	HEPG2, HEK29T, HCT116, MCF7	HUVEC
Cell density per well	400 cells	636 cells (other cell densities on same plate not shown)	375 cells
Nb of pooled samples per plate	1536	1536 (32 replicates per cell type and density)	384 (alternate columns & rows used)

Table 1. Experimental details for each independent experiment.

Results

Reproducible high-throughput transcriptomics across experiments

To assess read depth uniformity across all wells, we first prepared libraries from a full 1536-well plate containing 400 HEPG2 cells per well. After processing the plates and sequencing, samples were demultiplexed using their unique well-specific barcodes. More than 96% of reads were successfully demultiplexed and assigned to their correct sample of origin, indicating the barcoding chemistry is robust (Fig. 2A). Similarly, an independent experiment performed on a quadrant of a 1536-well plate (n=384) with 375 HUVECs per well also achieved a demultiplexing rate of over 96%, independently confirming the sample barcoding approach used (Fig. 2A).

Next, we assessed plate-wide uniformity to determine any position-related effects. The full 1536-well HEPG2 experiment had largely uniform depth across the plate, averaging around 180K reads per sample, with only a minor edge effect observed (Fig. 2B). This indicates that all 1536 wells can be effectively used across high-throughput screening studies, maximizing experimental output. The majority of these reads uniquely mapped to exons or other genomic features in both experiments (HEPG2, 82.6%; HUVEC, 85.7%), again indicating reproducible high-throughput transcriptomics across experiments and cell types (Fig. 2C).

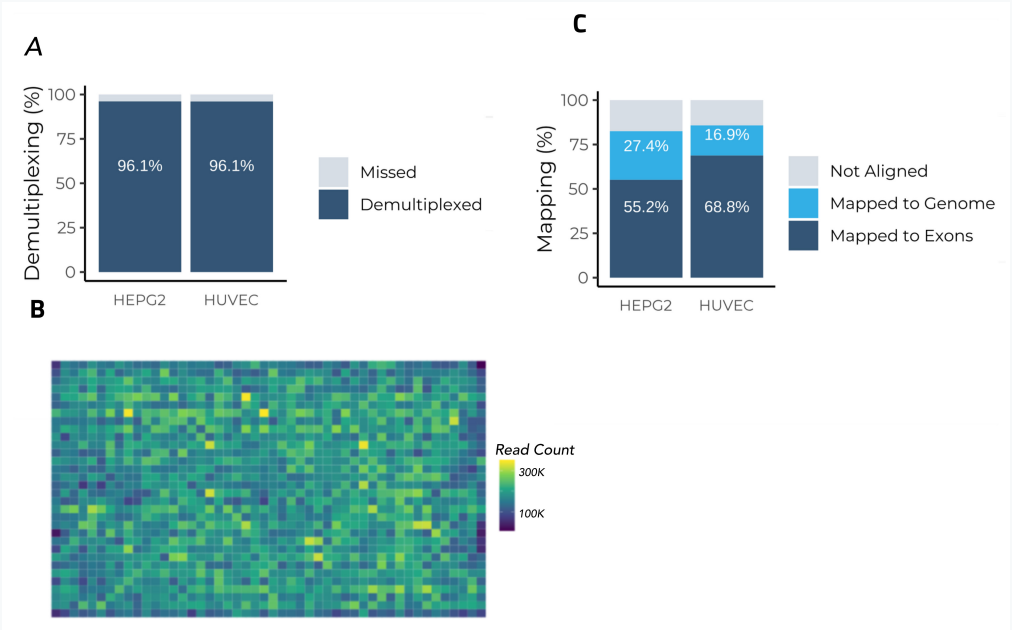


Figure 2. MERCURIUS™ 1536 DRUG-seq quality control and mapping statistics across two independent laboratories and cell types. (A) Demultiplexing rates for 1536-well plates containing HEPG2 cells and HUVEC cells. (B) Per-well read count heatmap across the full 1536-well plate from the HEPG2 experiment. (C) Read mapping statistics for both platforms, showing the proportion of reads mapping to exons, the genome, or unaligned for each experiment.

Independent Confirmation in Three Other Cell Lines

To assess the performance of MERCURIUS™ 1536 DRUG-seq in three other cell lines and to independently confirm the results previously observed for HEPG2 cells, we next prepared a 1536-well plate containing HEPG2, HEK293, MCF7, and HCT116 cells at different cell densities. Quality control metrics were broadly similar to those of the previous experiments for all cell lines. We focused on the 636-cells-per-well density, in line with the cell density recommended in the MERCURIUS™ 1536 DRUG-seq user guidelines (400-1,000 cells per well). For each cell line, between 9,000-10,000 genes were detected at only 200,000 reads per sample (Fig. 4A). This gene detection rate remained largely stable regardless of varying read depth per sample (Fig. 4B).

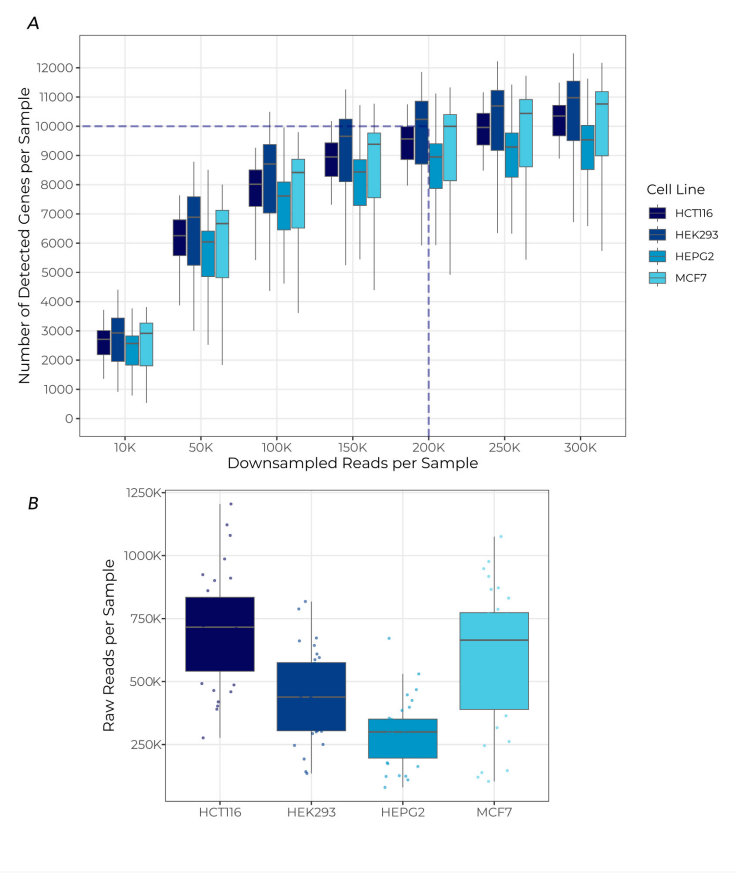
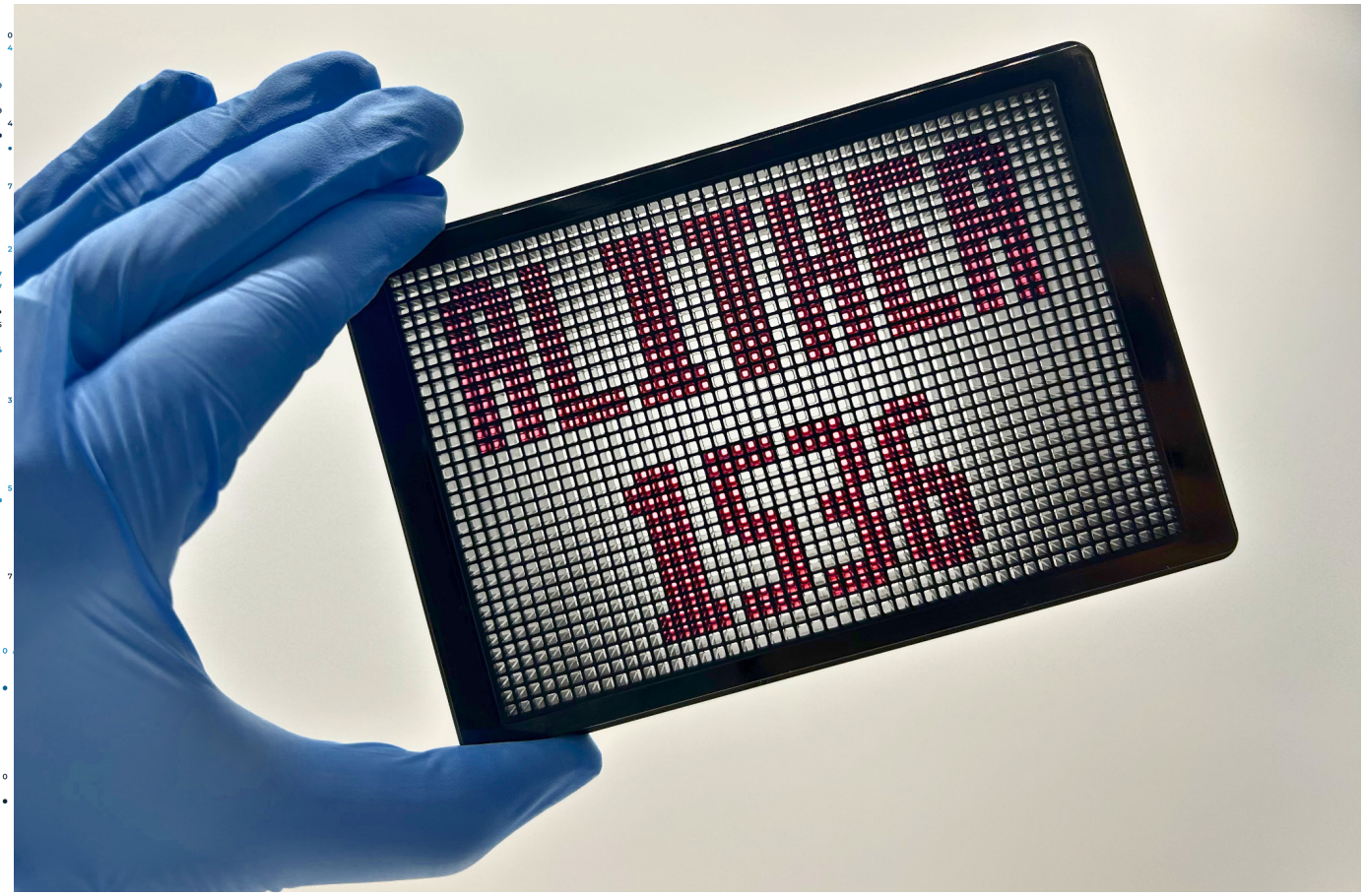


Figure 4. Gene detection rate and raw reads per sample across four cell lines.
(A) Gene detection saturation curves for HCT116, HEK293, HEPG2, and MCF7 cell lines, showing the number of detected genes per sample versus downsampled read depth. Dashed lines indicate the gene detection rate at a very low downsampled sequencing depth of 200K reads per sample.
(B) Raw reads per sample for each cell line prior to downsampling, shown as box plots with individual sample values as dots.



The road to the million-sample atlas

Using 1536-well plates dramatically increases MERCURIUS™ DRUG-seq throughput by generating substantially more profiles per processed plate.
Scaling to this level of throughput with 384-well plates typically requires additional liquid-handling robots, expanded laboratory infrastructure, and a larger operational footprint. In contrast, 1536-well plates significantly reduce the space and equipment required to process the same number of samples, enabling far greater screening efficiency (Table 1).
This increased density unlocks the generation of truly AI-scale datasets, reducing the time required to generate 1 million profiles to approximately three months and paving the way for screens an order of magnitude larger at a competitive cost.

	100,000 Profiles	1,000,000 Profiles
384WP – 20 plates/week	3.25 months	2.5 years
1536WP – 20 plates/week	<1 month	8 months
384WP – 50 plates/week	1.5 month	1 year
1536WP – 50 plates/ week	<2 weeks	3.25 months

Table 2: Estimated processing time for large scale drug screening

Conclusions

These three independent experiments across five different cell types demonstrate that MERCURIUS™ 1536 DRUG-seq delivers reliable, consistent transcriptome-wide profiling at screening scale, even at shallow sequencing depths. Demultiplexing rates consistently exceeding 96%, detection of over 10,000 genes at 200K reads per sample, and per-sample costs aligned with other screening technologies now establish the technical foundation for generating AI-scale transcriptomic datasets prospectively and at a throughput previously unattainable with sample-by-sample bulk RNA-seq or probe-based methods such as L1000.

MERCURIUS™ 1536 DRUG-seq represents a necessary step towards enabling biopharma to generate foundational, standardized, and proprietary AI training data moats directly on biologically relevant material at the scale required for more robust in silico models and more confident drug discovery and development. It now establishes a roadmap for generating millions of transcriptome-wide profiles as readouts of compound and dosage effects, impacting broad aspects of the biopharma pipeline.

References

1. Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, Li X, Wu JC, Yang S. Artificial intelligence in drug development. Nature Medicine. 2025 Jan;31(1):45-59.
2. Ye C, Ho DJ, Neri M, Yang C, Kulkarni T, Randhawa R, Henault M, Mostacci N, Farmer P, Renner S, Ihry R. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nature Communications. 2018 Oct 17;9(1):4307.
3. Li J, Ho DJ, Henault M, Yang C, Neri M, Ge R, Renner S, Mansur L, Lindeman A, Kelly B, Tumkaya T. DRUG-seq provides unbiased biological activity readouts for neuroscience drug discovery. ACS Chemical Biology. 2022 May 4;17(6):1401-14.
4. Way GP, Natoli T, Adeboye A, Litichevskiy L, Yang A, Lu X, Caicedo JC, Cimini BA, Karhohs K, Logan DJ, Rohban MH. Morphology and gene expression profiling provide complementary information for mapping cell state. Cell Systems. 2022 Nov 16;13(11):911-23.
5. Fredin Haslum J, Lardeau CH, Karlsson J, Turkki R, Leuchowius KJ, Smith K, Müllers E. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. Nature Communications. 2024 Apr 24;15(1):3470.
6. Ha SV, Jaensch S, Kańduła MM, Herman D, Czodrowski P, Ceulemans H. Cross modality learning of cell painting and transcriptomics data improves mechanism of action clustering and bioactivity modelling. Scientific Reports. 2025 Jul 2;15(1):23010.
7. Haghighi M, Caicedo JC, Cimini BA, Carpenter AE, Singh S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nature Methods. 2022 Dec;19(12):1550-7.