Element Biosciences

Correlating Teton™ with RNA–seq
Benchmarking Teton CytoProfiling on AVITI24™ with Bulk RNA–seq

Highlights

- Teton CytoProfiling delivers RNA expression data that exhibits high concordance with bulk RNA–seq, while capturing protein, morphology, and spatial context in a single assay.

- Differential expression results mirror bulk RNA–seq, enabling confident biological interpretation with richer cellular insight.

- Robust and reproducible performance across wells demonstrates reliability of Teton CytoProfiling on AVITI24.

Introduction

Bulk RNA–seq is a widely used and highly valuable method for measuring gene expression across diverse research and clinical applications, including cancer biology, and biomarker and gene fusion discovery[1]. However, because bulk RNA–seq averages expression across pooled cell populations, it cannot resolve cell-to-cell heterogeneity, spatial context, or cell morphology, and may miss signals from rare subpopulations of cells[1].

Teton CytoProfiling on AVITI24 enables multimodal profiling of RNA and protein expression alongside cell morphology and spatial organization at subcellular resolution in a single assay. The integrated approach of Teton CytoProfiling provides a more complete view of cellular biology, while maintaining compatibility with established gene expression workflows.

Here, we benchmark Teton CytoProfiling RNA measurements against the gold-standard bulk RNA–seq using matched datasets from three human cell lines (HeLa, A549, and HUVEC). By evaluating gene-level expression and biological concordance, we demonstrate that despite methodological differences, both approaches support concordant biological conclusions.

Teton CytoProfiling delivers RNA expression data that is consistent with bulk RNA–seq, while providing rich cellular

context and deeper biological insight[2]. In contrast, bulk RNA–seq measures pooled transcripts and lacks subcellular spatial context because cells are lysed prior to analysis.

Experimental Design

Bulk RNA–seq and Teton CytoProfiling data was generated from three human cell lines: HeLa, A549, and HUVEC. A Teton CytoProfiling run on AVITI24 for each cell line was performed using the Teton Human MAPK–Cell Cycle Panel Kit which includes 350 RNA targets, 50 protein targets, and 6 cell paint markers. A single bulk RNA–seq run for all three cell lines was performed on AVITI24. See Sample Preparation Methods section for details on cell culture, library prep, Teton CytoProfiling assay, and sequencing.

An overlapping subset of genes from both datasets was used to correlate the RNA expression data from Teton and bulk RNA–seq. A total of 339 genes were analyzed from the Teton and bulk RNA–seq datasets to produce the correlations. See Data Analysis Methods section for data processing and analysis details.

Adjustment by Probe Concentration

Teton CytoProfiling RNA is measured using a panel of oligonucleotide probes mixed at different relative concentrations based on transcript expression levels to mitigate signal saturation[2]. Adjusting the Teton datasets according to batch and probe concentration is an essential step in data processing before making any correlations between datasets. These adjustments account for any technical variability between batches and improve gene-to-gene comparability because probe concentrations differ across targets in the panel. The probe concentrations differ to account for differences in their expression including high expressors like housekeeping genes. See Data Analysis Methods section for details.

Impact of Normalization on Correlation

Normalization strategy influences correlation between Teton and bulk RNA–seq data by affecting the contribution of low–

abundance transcripts and accounting for technical variability. After adjusting Teton counts for probe concentration, both datasets were normalized to enable cross-platform comparison.

Distinct normalization approaches were applied to reflect assay-specific measurement characteristics. Bulk RNA-seq measures fragmented transcripts so data was normalized using transcripts per million (TPM) to account for gene length and sequencing depth. Because Teton measures individual RNA molecules, only depth adjustment is required. Accordingly, Teton data for each cell line was normalized using counts per 10,000 detected molecules (CP10K), a strategy suited to lower count, plexity, and variability between cells[3], similar to single-cell RNA-seq workflows.

Teton data can also be processed using single-cell–specific normalization strategies (see Drug-Induced Liver Injury and T Cell Activation Data Spotlights for details). However, for this study both datasets were treated as bulk to ensure consistency in correlation analysis.

Sequencing and Cytoprofiling QC

Standard sequencing and cytoprofiling QC metrics confirmed that both assays performed as expected. The bulk RNA-seq run sequenced with Cloudbreak Freestyle™ sequencing workflow produced ~1.4 B reads with good quality (over 94% Q30).

Teton Dataset Well-to-Well Correlation

Each Teton flow cell contained 12 wells with replicates from the same cell line. To check for reproducibility across the flow cell, well-to-well gene count correlations within each Teton CytoProfiling run were assessed. Pearson correlations of 0.992–0.997 indicate very high well-to-well correlations (Figure 1) highlighting the robustness and reproducibility of the results across each individual Teton flow cell.

Teton RNA Data Correlates with Bulk RNA-seq

All correlations were conducted after probe adjustment, normalization, and log transformation. Correlations between the RNA data from the Teton and bulk RNA-seq runs are shown in Figure 2 for each cell line. While a perfect correlation is not expected between these two distinct assays, the Teton and bulk RNA-seq data showed strong positive correlation across all cell lines. Pearson correlation coefficients were 0.835 for HeLa, 0.836 for A549, and 0.722 for HUVEC.
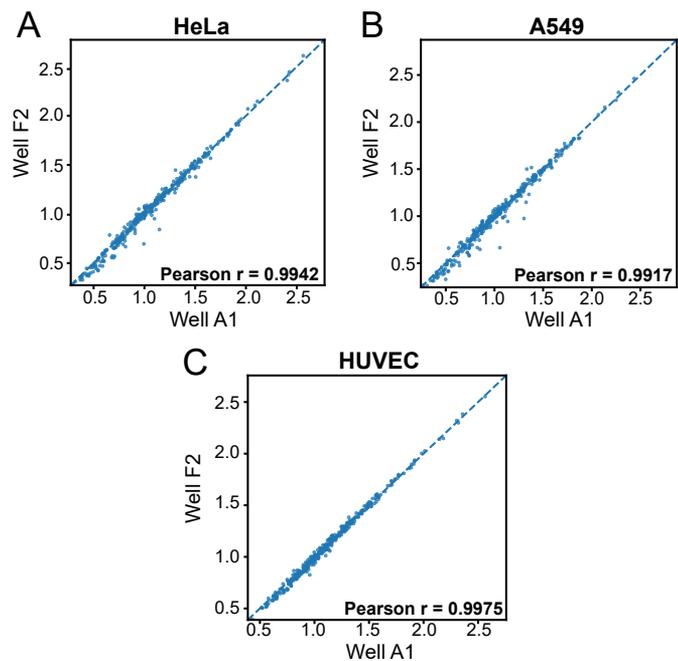

Figure 1. Well-to-well correlations for each cell line: (A) HeLa, (B) A549, and (C) HUVEC. One 12-well flow cell per cell line showing wells A1 vs. F2 for each cell line. All data was normalized using counts per 10,000 molecules detected (CP10K) and log-transformed ($\log_{10}$(CP10K+1)) before correlation.
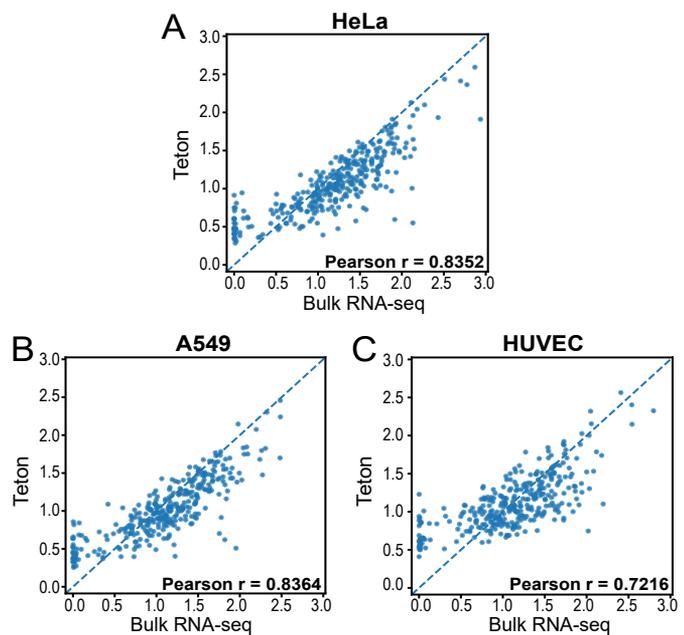

Figure 2. RNA data correlation between Teton and bulk RNA-seq for each cell line: (A) HeLa, (B) A549, and (C) HUVEC. Each point represents a unique transcript included in the Teton Human MAPK-Cell Cycle panel. Teton data was normalized using counts per 10,000 molecules detected (CP10K) and log-transformed ($\log_{10}$(CP10K+1)), and bulk RNA-seq data was normalized using transcripts per million counts (TPM) and log-transformed ($\log_{10}$(TPM+1)) before correlation.

Correlating Teton with RNA-seq Technical Note

For Research Use Only. Not for use in diagnostic procedures. | LT-00069 Rev. A | March 2026

2

Concordant Differential Expression

Differential expression analysis demonstrated strong concordance between Teton and bulk RNA-seq data across the pairwise comparisons of HeLa, A549, and HUVEC cell lines (Figure 3). Because these comparisons were performed between untreated cells at baseline, widespread differential expression across the MAPK-focused panel was not expected. In this setting, differential expression analysis is used primarily to assess cross-platform concordance and biological consistency rather than to capture a treatment-induced transcriptional response.

For each comparison, differential expression was calculated as the difference between the $\log_{10}$-transformed expression values between cell lines within each platform, using genes detected in both datasets and filtered to retain top differentially expressed genes and exclude lowly expressed genes (see Data Analysis Methods for details).

In the HeLa vs A549 comparison, 111 differentially expressed genes met inclusion criteria, yielding a Pearson correlation of 0.91 between platforms, with 97.3% of genes showing consistent differential expression directionality (Figure 3A). The HUVEC vs HeLa comparison included 119 genes and demonstrated a Pearson correlation of 0.81 with 84% directional agreement (Figure 3B). The HUVEC vs A549 comparison included 214 genes, with a Pearson correlation of 0.81 and 85.3% directional agreement (Figure 3C). Together, these results indicate a robust concordance between Teton and bulk RNA-seq in identifying differentially expressed genes across cell line comparisons.

Analysis of the top differentially expressed genes between cell lines further supports the biological relevance of the Teton assay. CDKN2A is a tumor suppressor and marker of epithelial cancer states[4]. CDKN2A was upregulated in the HeLa cell line compared to the HUVEC and A549 cell lines (Figure 3A, 3B). Endothelial markers (KDR and FLT1[5]) were upregulated in HUVEC compared to the HeLa and A549 cell lines, as expected (Figure 3B, 3C).

Clear relative expression differences between cell lines in the Teton dataset, together with consistent expression directionality compared to bulk RNA-seq, demonstrates that Teton CytoProfiling captures meaningful biological variability. Directional agreement between the Teton and bulk RNA-seq datasets is high across cell line comparisons. As expected, bulk RNA-seq shows larger differences of log-transformed expression values due to its greater sequencing depth. Teton shows the same directional trends as bulk RNA-seq but with a smaller dynamic range, since fewer molecules contribute to each gene's pseudo-bulk estimate. Despite this difference in dynamic range, Teton demonstrates strong concordance with bulk RNA-seq in identifying differentially expressed genes.

Summary

Bulk RNA-seq remains a gold standard for gene expression

analysis but it provides transcriptome data with limited biological context. Teton CytoProfiling delivers RNA expression data that closely correlates with bulk RNA-seq, supporting comparable analyses and biological conclusions.

In a single assay, Teton simultaneously captures RNA and protein expression, cellular morphology, spatial organization, and subcellular localization providing a true multiomic, spatially resolved view of the cellular state as compared to bulk RNA-seq.
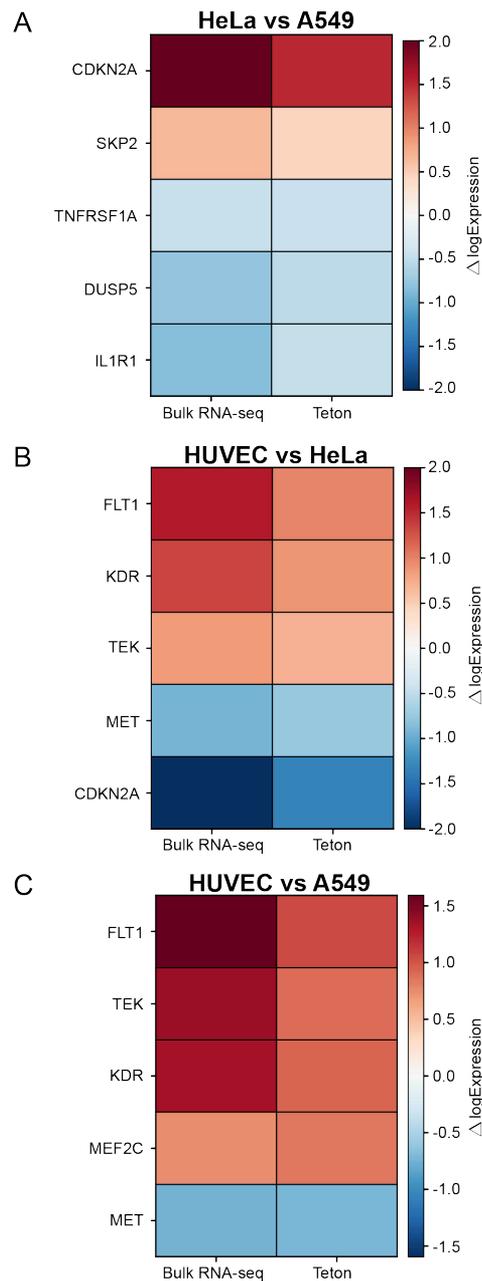

Figure 3. Selection of differentially expressed genes in pairwise cell line comparisons. Difference between the log-transformed expression values is shown for each gene from bulk RNA-seq (left) and Teton (right) datasets.

Correlating Teton with RNA-seq Technical Note

For Research Use Only. Not for use in diagnostic procedures. | LT-00069 Rev. A | March 2026

3

Sample Preparation Methods

Cell Culture

HeLa and A549 cell lines were obtained from the American Type Culture Collection (ATCC), and the HUVEC line was obtained from PromoCell. All cell lines were grown and maintained under standard growth conditions.

The culture media for the HeLa cells included Dulbecco's Modified Eagle Medium supplemented with heat-inactivated fetal bovine serum (FBS), penicillin-streptomycin, sodium pyruvate, and MEM non-essential amino acids. A549 cells were cultured in F-12K Nutrient Mixture supplemented with FBS and penicillin-streptomycin. The culture media for HUVEC included M199, Endothelial Cell Growth Medium (ECGM), MEM non-essential amino acids, sodium pyruvate, FBS, and penicillin-streptomycin. All cell lines were passaged every 2–3 days, seeding a T-75 flask with 500,000 cells for each respective cell line.

Bulk RNA–seq Library Prep

Live cells were washed, lysed, and RNA was extracted with the Direct-zol RNA kit (Zymo Research), following the manufacturer's protocol. For each cell type, 500,000 cells were extracted in 600 µL of TriZol reagent.

After RNA extraction, 200 ng of RNA from each sample was used as input into the Watchmaker RNA Library Prep Kit with Polaris Depletion, following the manufacturer's protocol. RNA-seq libraries were prepared with xGen Stubby Adapter-UDI Primers for Element (IDT) and were amplified with 10 cycles of PCR. RNA libraries were pooled and sequenced on AVITI24 with an Element 2×75 Cloudbreak Freestyle High Output Sequencing Kit.

Teton CytoProfiling Assay

Before seeding, cells were dissociated, moved to fresh media, counted with a hemocytometer, and diluted for the desired seeding concentration. Each cell line was seeded onto one 12-well, PLL-coated Teton slide and cells were incubated at 37°C with 5% CO_2 for 16–24 hours. HeLa cells were seeded at 9,000 cells per well, A549 cells were seeded at 9,000 cells per well, and HUVEC were seeded at 12,000 cells per well.

After the desired confluency was achieved, cells were washed and fixed, and flow cells were assembled according to the Teton CytoProfiling User Guide. Each flow cell was run on an AVITI24 using the Teton Human MAPK–Cell Cycle Panel Kit.

Data Analysis Methods

Bulk RNA–seq Data Processing

After sequencing, the RNA-seq data was demultiplexed with Bases2Fastq. The resulting FASTQ files were aligned with the splice-aware aligner STAR[6] and gene counts and normalization were both carried out in Salmon[7]. Transcripts per million (TPM) counts were output by Salmon and were log-transformed before correlation with the Teton data.

Teton CytoProfiling Data Processing

Teton data processing for correlation analysis consists of importing the RawCellStats output file, filtering, target selection, normalization, adjusting probe concentration, computing data, removing non-target features, aggregation, and log transformation.

The RawCellStats.parquet output generated after a Teton CytoProfiling run contains raw counts for all measured Teton features, including RNA and protein targets, morphology metrics, and assay controls. Each row in the file corresponds to a single cell and each column represents total counts per target. A gene count represents a detected molecule for that target.

Batches comprise groups of targets and assay controls sequenced and imaged together to prevent optical crowding. The assay controls help in assessing assay performance and include:

- **Non-specific binding (NSB) probes:** Four probes per batch designed to target no known transcripts to establish a noise floor
- **GAPDH probes:** Included as a positive control across multiple batches at low probe concentration

Filtering and Target Selection

The RawCellStats file was filtered using the *filter_cells* command in the Cytoprofiling package to remove poor-quality cells and segmentation artifacts, ensuring robust downstream analysis. After filtering, a total of 60 K HeLa cells, 126 K A549 cells, and 68 K HUVEC were used for analysis.

A subset of Teton RNA targets were then selected from the file to match targets from the bulk RNA-seq TSV file using Ensembl gene IDs or gene symbols. Only RNA targets shared between the Teton and bulk RNA-seq datasets were used for analysis.

Probe Concentration Adjustment

Counts were adjusted by batch using the Cytoprofiling package to account for any technical variability between each set of targets as batches are separated by time. Relative per-gene probe concentrations were extracted from the panel.json file. The maximum finite, non-zero concentration in the panel was used as a reference concentration. For each gene, a scaling factor was calculated as the ratio of the reference concentration to the gene-specific probe concentration. Raw per-cell counts were multiplied by the corresponding gene-specific scaling factor to generate a probe-concentration–normalized count matrix.

Normalization

Prior to correlation with bulk RNA-seq, the Teton dataset was normalized. Teton counts were normalized to counts per 10,000 molecules detected to account for Teton assay plexity

and avoid inflating or compressing the dataset.

Removing Non-target Features

Metadata and non-target features were removed from the RawCellStats file. This included NSB and unassigned counts (reads not confidently mapped to any target in the Teton Human MAPK-Cell Cycle Panel) as they do not correspond to biological RNA targets used in downstream analyses. Multiple GAPDH columns (present across batches) were averaged to generate a single GAPDH measurement per cell.

Pseudobulking and Well-to-Well Correlation

To enable comparison with the bulk RNA-seq, Teton data was pseudobulked by well (for well-to-well reproducibility) or by condition/cell line.

For well-to-well correlations, raw counts were summed per gene per well, then counts for each gene were divided by 10,000 cells per condition, and log-transformed. Pearson correlations were performed between wells to assess reproducibility across each flow cell. For each run, wells A1 and F2 were selected for comparison as they are positioned furthest apart on the flow cell and were most likely to present the greatest variation.

For bulk-seq RNA comparisons, Teton data was pseudobulked by condition (same cell line) by summing raw counts per gene across all wells corresponding to the same condition. Then each gene's counts were divided by 10,000 cells per condition and log transformed.

Pearson Correlation

Pearson correlations between the Teton runs and bulk RNA-seq TPM were performed after the data from both assays were normalized and log-transformed. All correlations were unweighted with no additional gene filtering performed at this stage.

Bulk RNA-seq and Teton expression data was compared using a gene-wise correlation. Expression values from both datasets were combined into a single gene-by-sample matrix. Expression concordance was quantified using a Pearson correlation (from pandas built-in implementation in Python) computed across genes for each pair of samples (Teton and bulk RNA-seq).

Differential Expression Analysis

Differential expression concordance between Teton and bulk RNA-seq was evaluated across pairwise comparisons of HUVEC, HeLa, and A549 cell lines. For each platform, differential expression was calculated as the difference in log-transformed expression values between cell lines (e.g., HeLa minus A549). These matched differences were used as comparable gene-level differential expression values across platforms.

Concordance was assessed using two metrics:

- Pearson correlation between matched differential expression values across shared genes.
- Directional agreement which is defined as the fraction of genes with matching differential expression direction (positive or negative) between Teton and bulk RNA-seq. Only genes where the absolute differential expression exceeded 0.2 in either platform were considered. This threshold was applied to focus the analysis on strongly differentially expressed genes and exclude genes with negligible cell line differences.

Relevant application datasets can be downloaded at Element Biosciences Datasets.

To learn more, visit elementbiosciences.com

References

1. Li, X. & Wang, C.-Y. From bulk, single-cell to spatial RNA sequencing. Int J Oral Sci 13, 36 (2021).

2. Lopez, T. et al. High-Throughput Multiomics Profiling of Model Systems Using the AVITI24 Platform. (May 2025): https://doi.org/10.1101/2025.05.03.651997.

3. Stegle, O., et al. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16, 133–145 (2015).

4. Liggett, W. H. & Sidransky, D. Role of the p16 tumor suppressor gene in cancer. J Clin Oncol 16, 1197–1206 (1998).

5. Yang, S. et al. Vascular Endothelial Growth Factor–Induced Genes in Human Umbilical Vein Endothelial Cells: Relative Roles of KDR and Flt-1 Receptors. ATVB 22, 1797–1803 (2002).

6. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).

7. Patro, R., et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419 (2017).

Element Biosciences

Telephone: 619.353.0300
Email: info@elembio.com

Correlating Teton with RNA-seq Technical Note

For Research Use Only. Not for use in diagnostic procedures. | LT-00069 Rev. A | March 2026

5