Sensitive Gene Expression Profiling with TempO-Seq: Advantages and Applications

Introduction

TempO-Seq (Templated Oligo assay with Sequencing readout) is a high throughput targeted sequencing assay, enabling comprehensive transcriptome analysis (1). It measures specific sequences within each gene while assessing every gene directly from RNA in crude lysates or purified RNA. This is achieved through the hybridization and ligation of detector oligos (DOs) to target sequences, followed by enzymatic removal of excess DOs. The ligated DO pairs are then amplified and sequenced, offering high specificity even for highly similar genes (Fig. 1). TempO-Seq's biochemistry significantly reduces mis-ligation and background noise, allowing precise measurements without bias, whether from bulk samples or single cells. Ligation provides single base specificity to measure expressed variants as well as differentiate highly homologous genes within or between species. Signals from highly abundant transcripts can be specifically and quantitatively attenuated to utilize sequencing space more efficiently and increase sensitivity to measure low expressed genes, without impacting differential expression analysis. DOs can target specific splice junctions within a gene, permitting differential measurement of alternatively spliced isoforms.

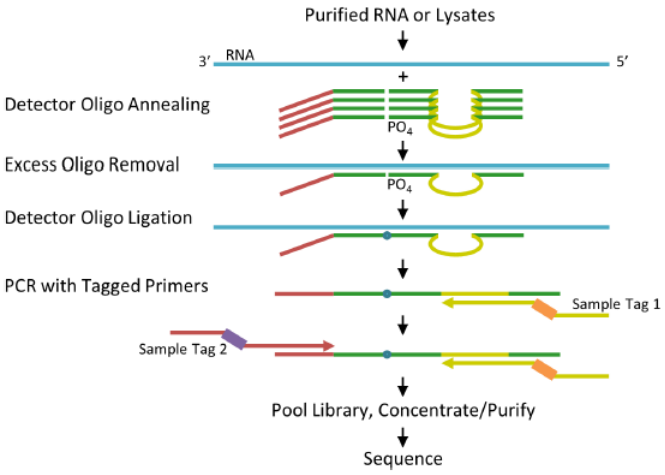

Fig 1. TempO-Seq biochemical scheme. RNAs are targeted by annealing to DOs that contain target specific sequences (green) as well as primer landing sites (red and yellow) that are shared across all DOs. Excess oligos are removed by an exonuclease, then the hybridized oligos are ligated and amplified using primers that contain sample tag (index) sequences (orange and purple bars), and adaptors required for sequencing. The amplified assay products are pooled into a library, purified, concentrated and sequenced.

TempO-Seq presents several advantages over RNA-Seq and other targeted sequencing methods for gene expression profiling whether it is to measure the whole transcriptome or targeted sets of genes. Unlike RNA-Seq, TempO-Seq eliminates the need for time-consuming RNA purification, quantitation, and cDNA synthesis steps, reducing variability and increasing sensitivity. Quantitative attenuation, referred to above, is not possible using RNA-seq. A differentiating hallmark of TempO-Seq is the reproducibility of measurements. Technical replicates for the whole transcriptome correlate with an $R^2 > 0.97$ (1). Since it is the ligated synthetic DO adducts that are sequenced, and the sequence of each is known, alignment and data analysis to generate a table of gene expression levels in each sample is simplified and data analysis can be easily standardized and automated to generate the sample gene expression table. These features make TempO-Seq an ideal transcriptomics platform for discovering new gene associations and signatures for diseases or treatments. With an innovative enrichment strategy, TempO-Seq is also used to detect low level pathogens (RNA or DNA) in dilute samples in an ultra-high throughput fashion.

Table 1. Comparison of key features for TempO-Seq and RNA-seq for human whole transcriptome analysis

Human whole transcriptome		TempO-Seq	RNA-Seq
Input sample type	Purified RNA	1-100ng	10-1,000ng
	Crude Cell Lysate	Good	N/A
	FFPE	Good	Poor
cDNA synthesis		No	Yes
3' bias		No	Often
de novo		No	Yes
Attenuating abundant transcripts		Yes	No
Sequencing reads per sample		4-5M	15-30M
Data Analysis		Simple & Fast	Complicated & Slow

Maximizing Sequencing Output through Attenuation of Highly Abundant Transcripts

To use sequencing reads efficiently and increase the sensitivity to measure low abundant genes, the signal from high abundance transcripts in the TempO-Seq assay can be attenuated by introducing non-functional competitor oligonucleotides at a chosen ratio per transcript. This attenuation was designed to quantitatively reduce the signal from highly abundant targets without eliminating them entirely. Read counts for attenuated genes can be corrected by multiplying them by the ratio of attenuator to DO for each gene to calculate their unattenuated read counts. Such back-calculated data correlates with unattenuated data, demonstrating its validity (Fig. 2).

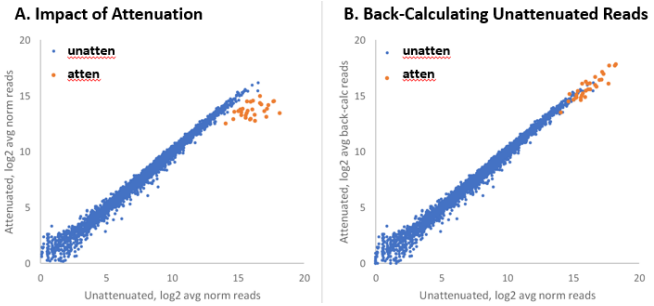

Fig 2. Back-calculation of attenuated genes. (**A**) Samples were assayed with or without attenuation and compared. The attenuated genes are highlighted in orange, showing that their read counts are lowered. (**B**) The attenuated targets are back calculated by multiplying by fold attenuation for each target. The back-calculated read counts (orange) are in alignment with the unattenuated population (blue).

Measuring Each Gene Independently Without Impact on Other Genes When Changing Content

Unlike arrays, each gene measure is independent of the measurement of every other gene in the assay. Thus, whole transcriptome data can be used to implement a focused assay of only the genes of interest without change in results, or content can be added to an existing assay again without changing results (1). This is illustrated by Fig. 3 which demonstrates that the measurements of a set of 1,396 genes that are in common between two different assays (measuring either 967 or 1,545 additional genes, respectively) are not impacted by the other genes measured in each assay and give the same expression count data (R^2 = 0.98). Thus, content can be added to any of the commercial off-the-shelf TempO-Seq assays, or any custom assay.

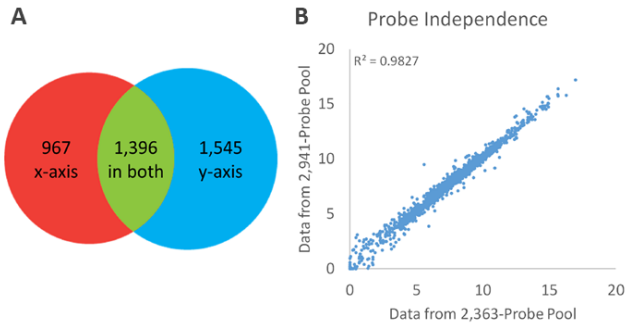

Fig 3. Context independence. (**A**) The overlap between two DO pools is illustrated. Red indicates the DOs only in the 2,363-plex pool, blue indicates the DOs only in the 2,941-plex pool, and green indicates the 1,396 DOs that are present in both pools. (**B**) Read counts for the 1,396 transcripts targeted in both pools are compared (R2 = 0.98).

Comparing TempO-Seq and RNA-Seq for Differential Gene Expression Analysis

The accuracy of TempO-Seq was verified for measurement of differential gene expression by comparing it to RNA-Seq data using RNA purified from MCF-7 and MDA MB 231 cells (1). The results showed a strong correlation (R^2 = 0.91) between the log fold changes detected by TempO-Seq and RNA-Seq, demonstrating consistent measurement of differentially expressed genes (Fig. 4). Additionally, using synthetic ERCC RNA Mixes, TempO-Seq accurately reported known fold differences (R^2 = 0.97), further confirming the quantitative accuracy in measuring differential gene expression.

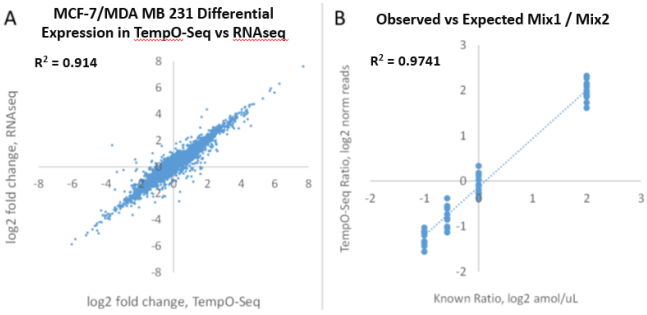

Fig 4. Cross-platform comparison of TempO-Seq and RNA-Seq. (**A**) Log2 fold changes between MCF-7 and MDA MB 231 RNAs as measured by TempO-Seq (~4 M reads/sample) are compared to those measured by RNA-Seq (~15 M reads/sample) for the 6,500 genes that had over 20 read counts in both cell types. R^2 = 0.91. (**B**) ERCC Mixes 1 and 2

were diluted 1:1,000 and spiked into 100ng Universal Reference RNA, then assayed with a dedicated DO pool. Fold differences in sequencing reads are compared to the fold differences in concentration between the two mixes. Read depth was 340K reads/sample. $R^2 = 0.97$.

Independent Comparison of TempO-Seq to Microarray and RNA-Seq

In a recent study (2), Bushel et al conducted a comprehensive comparison of the TempO-Seq platform with two well-established gene expression analysis methods: microarray and RNA-Seq. They used 45 purified RNA samples extracted from the livers of rats exposed to chemicals categorized into five different modes of action (MOAs). These samples had previously undergone assessment using Affymetrix™ rat genome 230 2.0 microarrays and Illumina® whole transcriptome RNA-Seq. The comparative analysis revealed some variations in signal-to-noise ratios, root mean squared error, and sources of variability between the different platforms. Notably, microarray and TempO-Seq exhibited the highest consistency in capturing variability related to MOAs and chemical treatments. In contrast, RNA-Seq exhibited more noise and greater differences between samples within the same MOA category.

An in-depth examination of the data using various analytical techniques, including hierarchical clustering, gene subnetwork connectivity, and representation of biological processes associated with MOA-varying genes, demonstrated that results from the TempO-Seq platform are consistent with findings on other more established approaches for measuring the genome-wide transcriptome (Fig. 5).

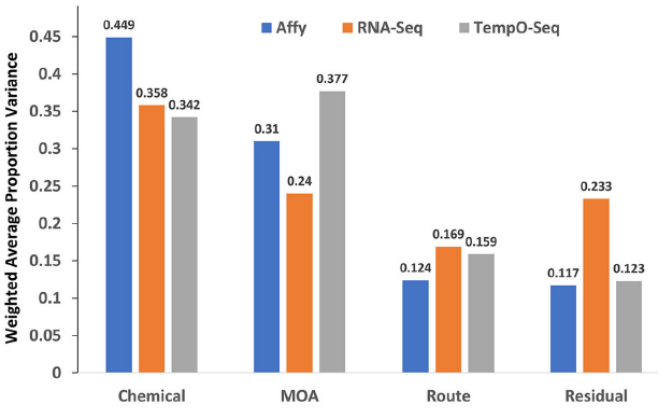

Fig 5. Variance components explained. Shown on the y-axis is the weighted average of the proportion of variance explained by platform for each of the mixed effect linear model terms denoted in the x-axis.

High Throughput Pipeline for Dose Response Toxicogenomic Modeling

TempO-Seq was used to profile cellular responses to Trichostatin A (TSA), monitoring gene expression changes in MCF-7 cells exposed to various TSA concentrations (1). TempO-Seq effectively detected gene expression changes at a TSA concentration of 1 µM, consistent with published data. Using DESeq2 analysis, 4,154 differentially expressed transcripts were identified as responsive to TSA compared to a control group. TempO-Seq results were compared with the Connectivity Map (cMAP) database, revealing overlapping gene sets related to TSA treatment across different cell types. TempO-Seq also uncovered novel TSA-responsive genes that cMAP missed, demonstrating its enhanced sensitivity, which was confirmed through cross-platform comparisons and RT-PCR. TempO-Seq identified a set of 1,571 genes affected by TSA across multiple cell types, shedding light on potential transcription-related mechanisms. Traditional toxicogenomics has mainly focused on individual gene concentration-responses of a limited number of substances, but TempO-Seq offers a cost-effective way to study the entire transcriptome, making it suitable for high-throughput analysis of drug and chemical concentration-response relationships. Multiple research groups have utilized this capability. This TempO-Seq enabled pipeline can be adapted for various types of chemical response studies (3), even when sample sizes are small. Examples include profiling the effects of different doses of four chemicals on gene expression in induced pluripotent stem cell cardiomyocytes (3), demonstrating the importance of determining the Point of Departure (POD) from dose-response relationships. The POD represents the concentration that significantly departs from the control mean, indicating the potential impact of a chemical on gene expression (Fig. 6). Dose response TempO-Seq data has been used to assess the differential safety of compounds with similar clinical efficacy (4), using an analysis pipeline referred to as BMDExpress to calculate benchmark doses, analogous to $EC_{50's}$ for every gene measured. TempO-Seq enables an automated framework for assessing gene expression responses to varying concentrations of drugs and chemicals. These responses can be valuable for hazard or efficacy assessment, drug repurposing, and disease characterization, providing

insights into the impact of different compounds on gene expression in a high-throughput manner.

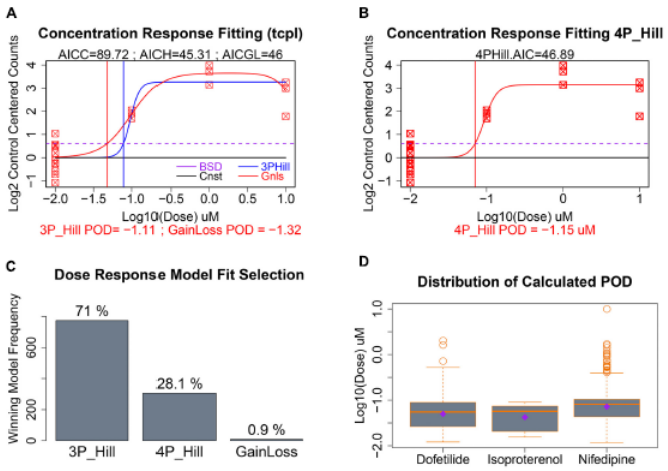

Fig 6. Concentration response modeling and POD calculation. (**A**) Baseline deviation (BSD) is set to 1 SD from the mean of controls (dotted purple line). The tcpl package is used to fit constant, gain loss, and 3-parameter hill models. (**B**) The DRC package in R is used to fit a 4-parameter hill function and the model with the smallest AIC is chosen as "best" with corresponding POD used where fitted model crosses BSD. (**C**) Distribution of the winning model for these data. (**D**) Evaluation of the dispersion, mean, and median of calculated POD deviations by chemical treatment.

Extraction-free Whole Transcriptome Gene Expression Analysis of FFPE Sections

TempO-Seq offers a groundbreaking advantage for gene expression profiling in formalin-fixed paraffin-embedded (FFPE) tissue samples. FFPE samples, which are commonly used in clinical pathology, are historically challenging to assess due to low RNA yield, fragmentation, and degradation during preservation and extraction. TempO-Seq's targeted, ligation-based approach makes it highly resistant to RNA fragmentation and provides quality data from RNA of RIN 3.0 (1), making it well-suited for FFPE samples, even those that are decades old, without the need for laborious RNA extraction. The sensitivity of TempO-Seq enables profiling of small focal areas of interest within FFPE tissues, and it can even be used on H&E stained tissues (5). TempO-Seq has been successfully used to detect differences in gene expression between normal and cancer samples highlighting its potential for identifying disease subtypes or molecular signatures among patients (Fig. 7). Overall, TempO-Seq's unique capabilities make it a significant advancement in the field of gene expression analysis from FFPE samples.

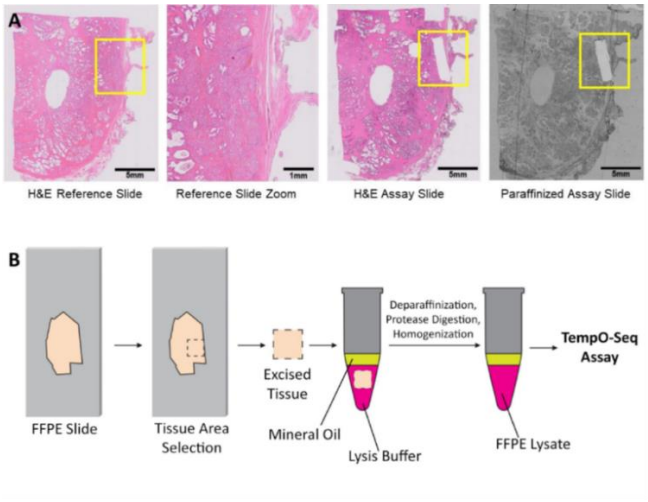

Fig 7. Processing FFPE samples for the TempO-Seq assay. (**A**) Examples of slides processed through the TempO-Seq FFPE assay. Left panel shows an H&E stained section, with the yellow box indicating the area of interest. Center left shows an expanded image of this area, demonstrating the mixed histology that would affect the data if the entire area were to be scraped and profiled. Areas identified from a stained tissue section can be scraped from an unstained, paraffinized adjacent section (right), or (if RNase-free reagents are used for staining) directly from a stained section (center right). The scraped areas in this case were approximately 1 x 5 mm, aligned with the focal histology of interest, and sufficient for gene expression profiling. (**B**) An area of interest is manually scraped from mounted FFPE sections. The tissue is added directly into 1X FFPE lysis buffer, overlaid with mineral oil, and then heated at 95℃ for 5 minutes. FFPE Protease is added and the sample is incubated and manually homogenized. The processed lysate is then ready for input directly into the annealing step of the TempO-Seq assay.

Summary

TempO-Seq is a fully automatable in-solution assay utilizing progressive dilution and addition-only techniques. With its targeted approach, and option to use attenuation, it demands less sequencing capacity, making it possible to conduct sensitive whole transcriptome profiling and alternative splicing analysis for large cohorts of samples. As a result, complex phenotypes and diseases demanding use of large numbers of samples for statistical robustness can employ TempO-Seq to delineate cellular, molecular, and disease or compound mechanisms with greater efficiency

and effectiveness than was previously achievable.

References

1. Yeakley JM, Shepard PJ, Goyena DE, et al. (2017) A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. **PLoS ONE** 12(5): e0178302. https://doi.org/10.1371/journal.pone.0178302
2. Bushel PR, Paules RS and Auerbach SS. (2018) A comparison of the TempO-Seq S1500+ Platform to RNA-Seq and Microarray Using Rat Liver Mode of Action Samples. **Front Genet** 9:485. doi: 10.3389/fgene.2018.00485
3. House JS, Grimm FA, Jima DD, et al. (2017) A Pipeline for High-Throughput Concentration Response Modeling of Gene Expression for Toxicogenomics. **Front Genet** 8:168. doi: 10.3389/fgene.2017.00168
4. Ramaiahgari SC, Auerbach SS, Saddler TO, et al. (2019) The power of resolution: Contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. **Toxicological Sciences** 169(2):553. doi:10.1093/toxsci/kfz065
5. Trejo CL, Babić M, Imler E, et al. (2019) Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue. **PLoS ONE** 14(2): e0212031. https://doi.org/10.1371/journal.pone.0212031